

R E P O R T   R E S U M E S

ED 002 772

24

AN APPLICATION OF REINFORCEMENT PRINCIPLES TO CLASSROOM TEACHING.

BY- PORTER, DOUGLAS

HARVARD UNIV., CAMBRIDGE, MASS., GRAD. SCH. OF EDUC.

REPORT NUMBER CRP-142

PUB DATE 31 MAY 61

HARVARD UNIV., CAMBRIDGE, MASS., LAB. RES. IN INSTR.

CONTRACT OEC-SAE-6899

EDRS PRICE MF-\$0.27 HC-\$5.60 140P.

DESCRIPTORS- \*TEACHING MACHINES, \*AUTOINSTRUCTIONAL AIDS, \*MECHANICAL TEACHING AIDS, \*ORTHOGRAPHIC SYMBOLS, \*PHONETICS, CLASSROOM TECHNIQUES, SPELLING, MASSACHUSETTS, CAMBRIDGE

THIS RESEARCH CONSIDERED THE TEACHING OF SPELLING AS A PHENOMENON MORE OR LESS ISOLATED FROM THE OTHER AREAS OF LANGUAGE ARTS INSTRUCTION THAT PROBABLY CONTRIBUTE TO SPELLING PERFORMANCE. THE MAJOR OBJECTIVES WERE-- (1) TO DESIGN, CONSTRUCT, AND TEST MECHANICAL DEVICES FOR THE TEACHING OF SPELLING, AND (2) TO CARRY OUT AN EXPLORATORY INVESTIGATION OF THE TEACHING DEVICES AND MATERIALS DEVELOPED IN ORDER TO DETERMINE THE FEASIBILITY OF MORE EXTENSIVE RESEARCH UTILIZING SIMILAR TECHNIQUES. THE TEACHING MACHINES DESIGNED FOR THIS RESEARCH PROVED TO BE SATISFACTORY AND DURABLE IN OPERATION, BUT LEFT MUCH TO BE DESIRED IN THE CONTROL OF CHEATING, SECURITY OF THE TEACHING MATERIALS, AND GENERAL CONVENIENCE. FOR RESEARCH PURPOSES, A FORM OF TEACHING MACHINE THAT ALLOWS SIMPLE PREPARATION OF MATERIALS AND EASE IN REVISION IS DESIRABLE. THE DESIGN OF SUCH A DEVICE IS SUGGESTED. OVERALL LEARNING OUTCOMES FROM THE TEACHING MACHINE TREATMENTS WERE GENERALLY COMPARABLE TO THOSE FROM THE USUAL CLASSROOM TECHNIQUES EMPLOYED IN THE EXPERIMENTAL SCHOOL, WITH SOME SUGGESTION OF SUPERIORITY FOR THE MACHINE TEACHING. WITHIN THE LIMITED FRAMEWORK OF THIS RESEARCH AN INTERESTING NEW TEACHING METHODOLOGY HAS BEEN TRIED WITH PROVOCATIVE RESULTS, AND SOME STEPS HAVE BEEN TAKEN TOWARD A BETTER UNDERSTANDING OF SPELLING AS A SUBJECT AREA. (GC)

ED002772

REFERENCE  
COPY

COOPERATIVE  
RESEARCH  
PROJECT

No. 142

A REPORT FROM THE LABORATORY  
FOR RESEARCH  
IN INSTRUCTION



GRADUATE SCHOOL  
OF EDUCATION

HARVARD  
UNIVERSITY



**U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE**

Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

**/AN APPLICATION OF REINFORCEMENT**

**PRINCIPLES TO CLASSROOM TEACHING: /**

**The teaching of phonemically and orthographically analyzed spelling materials by an automatic mechanical teaching device.**

**Principal Investigator**

**Douglas Porter**

**Laboratory for Research in Instruction**

**Graduate School of Education**

**Harvard University**

**Cambridge, Massachusetts**

**May 31, 1961**

**The research reported herein was performed pursuant to contract number SAE-6899 with the United State Office of Education, Department of Health, Education, and Welfare**

**COOPERATIVE RESEARCH  
PROJECT NO. 142**

PREFACE

The investigations reported in this monograph were carried out during the academic year 1957-58, before the field had become the object of public interest and financial speculation. The intervening period has seen increased activity by both psychologists and commercial sponsors, but with less accomplishment than one might expect. This is attested to by the fact that the contents of this report are still fresh,

In particular, no entirely satisfactory teaching machine for programming research has yet appeared on the market. Existing machines, some of them similar to the machines described in this report, have one or more flaws that make it difficult to control item presentation and arduous to prepare or revise the items. This is partly the result of psychologists trying to design machinery but more surely the result of trying to make hybrid machines suitable for both programming research and commercial use in schools. The flexibility and ease of item preparation and revision needed in a research device are not required for practical teaching. On the other hand, commercial devices should be designed for maximum reliability over the long run and use compact materials that can be mass produced easily. Better programs and more satisfactory practical results will come from recognition of these factors.

This work has been made possible by the generous cooperation of many individuals. I would like to thank especially Mr. Joseph C. Forte, Principal of the Mystic School in Winchester, Massachusetts, for making the facilities of his school available for an entire year of experimentation. Without the



close cooperation and assistance of his school staff this work would have been impossible: Miss Norma F. Littlefield and Miss Gladys L. Wood carried out the experimental programs in their classrooms, while Miss Martha Webb and Miss Eileen R. Cummings supervised the control groups; Mrs. Eleanor Geary and Miss Lydia J. Whittemore supervised some of the preliminary work in their classrooms for several months. I am also grateful to the School Secretary, Mrs. Maude G. Dover, for her constant help in procuring records, lists, and numerous other necessities.

At the Graduate School of Education, Mrs. Judith Burleigh Hulburt and Miss Barbara A. Durkee served as Research Assistants in the preparation of preliminary teaching materials and in testing and analysis of student performance. Mr. Albert E. Beaton, Jr. devised a machine analysis scheme that saved hours of hand tabulation of spelling words. Mrs. Nathalie Corse, who prepared all of the illustrations and cut all of the stencils on an ancient and balky Varityper, was absolutely indispensable. Miss Claudia P. Wilds contributed an excellent phonemicization. Corlis Knapp, Pearl Chan, and Marcia Mierzwa are to be credited with production of this report.

Finally, I would like to thank Dean Francis Keppel and Dr. John B. Carroll of the Graduate School of Education and Dr. B. F. Skinner of the Department of Psychology for their encouragement and assistance.

Cambridge, Massachusetts  
May, 1961

D.P.

# CONTENTS

|                                                                              | Page |
|------------------------------------------------------------------------------|------|
| PREFACE . . . . .                                                            | ii   |
| CONTENTS . . . . .                                                           | iv   |
| LIST OF TABLES . . . . .                                                     | v    |
| LIST OF FIGURES . . . . .                                                    | vii  |
| THE CONTRIBUTION OF TEACHING<br>MACHINES TO INSTRUCTIONAL RESEARCH . . . . . | 1    |
| DESIGN OF THE TEACHING DEVICE AND PROGRAMS . . . . .                         | 7    |
| A Simple Teaching Device . . . . .                                           | 7    |
| Redesign . . . . .                                                           | 13   |
| More Complex Teaching Devices . . . . .                                      | 18   |
| Fragments of a Spelling Program . . . . .                                    | 24   |
| Revision . . . . .                                                           | 30   |
| STUDIES IN THE ANALYSIS AND TEACHING OF SPELLING . . . . .                   | 34   |
| The Problems Studied . . . . .                                               | 34   |
| Teaching Efficiency . . . . .                                                | 35   |
| Year-long Changes in Performance . . . . .                                   | 37   |
| Individual Variables . . . . .                                               | 39   |
| Programming Factors . . . . .                                                | 42   |
| On English Spelling . . . . .                                                | 46   |
| Method . . . . .                                                             | 64   |
| Subjects . . . . .                                                           | 64   |
| Measures . . . . .                                                           | 66   |
| Procedure . . . . .                                                          | 68   |
| Results and Discussion . . . . .                                             | 71   |
| Teaching Efficiency . . . . .                                                | 71   |
| Year-long Changes in Performance . . . . .                                   | 80   |
| Individual Variables . . . . .                                               | 85   |
| Programming Factors . . . . .                                                | 94   |
| Transfer of Training . . . . .                                               | 101  |
| SUMMARY AND CONCLUSIONS . . . . .                                            | 107  |
| BIBLIOGRAPHY . . . . .                                                       | 114  |
| APPENDIX . . . . .                                                           | 117  |



# LIST OF TABLES

| Table |                                                                                                                                                                                      | Page |
|-------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1     | Variant Phonic Spellings of the Word "Fourteen" and Probability of the Spellings being Correct . . . . .                                                                             | 48   |
| 2     | Sequential Contingencies in the Spelling of /iy/ . . . . .                                                                                                                           | 51   |
| 3     | Percentage of Words Misspelled in Childrens' Letters by Grade in School . . . . .                                                                                                    | 58   |
| 4     | Characteristics of the 2nd, 4th, and 6th Grade Experimental and Control Groups . . . . .                                                                                             | 65   |
| 5     | Comparison of Median Grade Equivalent Scores of 2nd and 6th Grade Experimental and Control Groups before and after Treatment . . . . .                                               | 72   |
| 6     | Comparison of Median Spelling Gains of 2nd Grade Groups, Based upon Letter Scoring of the California Test . . . . .                                                                  | 75   |
| 7     | Number of Students per Week Finishing Lessons 23-34 on an <u>Ad Lib.</u> Basis . . . . .                                                                                             | 78   |
| 8     | Comparison of Mean Weekly Test Scores of 6th Grade Experimental Group for Different Parts of the Year . . . . .                                                                      | 83   |
| 9     | Comparison of Median Gains of the 6th Grade Groups Split into Upper and Lower Halves at the Median IQ Scores of Each . . . . .                                                       | 86   |
| 10    | Rank Order Correlations of Gains on the Stanford Test vs. IQ and Stanford Pretest Scores for Upper and Lower IQ Halves and Total 6th Grade Experimental and Control Groups . . . . . | 87   |
| 11    | Rank Order Correlations between Individual Spelling Gains on the Stanford Test and Age and Frequencies of Cheating, Misscoring, and Error . . . . .                                  | 91   |
| 12    | Rank Order Correlations of Average Weekly Test Scores for Individuals versus Age and Frequencies of Cheating, Misscoring, and Errors . . . . .                                       | 92   |
| 13    | Rank Order Correlations of the Average Frequency per Response of Cheating, Misscoring, Errors, and Responses per Lesson vs. Average Test Scores per Lesson . . . . .                 | 99   |

# LIST OF TABLES (continued)

| Table |                                                                                                                                                       | Page |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 14    | Comparison of 4th Grade Transfer (E) and Non-Transfer (C) Groups on Matching Variables . . . . .                                                      | 102  |
|       | The Teaching Device in Operation . . . . .                                                                                                            | 10   |
| 15    | Comparison of Average Percentage Spelling Errors in Written Work for 6th Grade Teaching Machine and Teacher Groups . . . . .                          | 104  |
| 1     | Sample of Student Handwriting . . . . .                                                                                                               | 22   |
| 5     | Year-long Changes in Performance of Students Taught by Teaching Machine . . . . .                                                                     | 81   |
| 6     | Average Incidence of Cheating, Misreading, and Errors for the 22 Machine Lessons Plotted as Average Frequency per Response $\times 10^{-3}$ . . . . . | 75   |



# THE CONTRIBUTION OF THE TEACHING MACHINE TO INSTRUCTIONAL RESEARCH

## LIST OF FIGURES

| Figure |                                                                                                                                                            | Page |
|--------|------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| 1      | Views of the Teaching Device . . . . .                                                                                                                     | 9    |
| 2      | The Teaching Device in Operation . . . . .                                                                                                                 | 10   |
| 3      | Redesigned Teaching Device . . . . .                                                                                                                       | 17   |
| 4      | Samples of Student Handwriting . . . . .                                                                                                                   | 22   |
| 5      | Year-long Changes in Performance of Students<br>Taught by Teaching Machine . . . . .                                                                       | 81   |
| 6      | Average Incidence of Cheating, Misscoring, and<br>Errors for the 22 Machine Lessons Plotted as<br>Average Frequencies per Response $\times 10^3$ . . . . . | 95   |

Throughout the world today there is a rising clamor for educational reform. The demand is for a new type of education, one that is more efficient, more economical, and more effective. The teaching machine has been hailed as the answer to this demand. It is a device that can deliver instruction in a systematic, self-paced manner. It can adapt to the individual needs of each student. It can provide immediate feedback. It can be used in a variety of settings. It can be used by students who are unable to attend school. It can be used by students who are unable to learn in a traditional classroom. It can be used by students who are unable to learn in a traditional classroom. It can be used by students who are unable to learn in a traditional classroom.

From the traditional position of academic isolation there has been a recent shift of interest by some psychologists toward the problems of education, especially the analysis of instruction. One aspect of this development that shows special promise is what has come to be known as programmed instruction. The basic conception is of carefully sequenced self-teaching

## THE CONTRIBUTION OF TEACHING MACHINES TO INSTRUCTIONAL RESEARCH

Throughout the world today there is a rising clamor for education from masses of people heretofore illiterate or only slightly educated. The call comes not only from underdeveloped areas of the world and not only for literacy, but also from the bastions of industrial society to improve existing education and to expand into ever-widening areas of specialized education. It is evident that this demand can be met adequately only through the development of a technology of education which will produce gains in efficiency, skill, and manpower economy of teaching. Historically, many contributions have already been made to this technology; the inventions of printing, motion pictures, and electronic mass media have all served to disseminate knowledge. But at the same time, these media have tended to decrease teacher-student contact and have deprived students of the evaluative interchange with a teacher which is a historically and, presumably, a psychologically important part of the teaching process. The research reported here represents a small part of the contemporary effort by psychologists to preserve the economic advantages of mass media and, at the same time, reinstate some of the advantages of face-to-face teaching methods.

From the traditional position of academic isolation there has been a recent shift of interest by some psychologists toward the problems of education, especially the analysis of instruction. One aspect of this development that shows special promise is what has come to be known as programmed instruction. The basic conception is of carefully sequenced self-teaching



materials that require explicit student responses and give immediate evaluation of performance. An adjunct to these teaching programs is the teaching machine, a device to ensure proper study of the programs by presenting teaching items in the desired order and by confirming student responses.

There are three historical antecedents to this development, one concerned with the planning and organization of subject-matter for teaching purposes, another concerned with the methodology of teaching, and a third derived from laboratory studies of learning. Teachers have always been more or less concerned with the organization of subject-matter into sequences designed to teach, but few formalized, tested, or generally accepted procedures or principles have been followed. The main guide-lines have come from the teacher's scholarly interest in the subject-matter. Educational psychologists and pedagogues, on the other hand, have been concerned with teaching methodology, but from a general point of view having little reference to the problems of teaching specific subject-matter. Psychologists engaged in the laboratory study of learning have generally occupied the role of scholar-teacher, concerned only informally with the problems of teaching. These three traditions have converged in the current work on teaching machines and programmed instruction.

In 1926, the educational psychologist Sidney L. Pressey described "A simple device which gives tests and scores--and teaches." Pressey was disturbed about the large amount of time teachers spent in routine clerical tasks, and designed a number of ingenious devices to lighten the load. He also found that the immediate self-scoring of tests made possible by his devices resulted in improved learning outcomes, previously discovered by Judd (1905-6),

Arps (1917), and others, and pointed out the advantages of having a student pace his own instruction. Unfortunately, Pressey's machines never caught on, but died the victims of cultural tradition and undernourishment, for Pressey and his associates had not vigorously attacked the problems of providing teaching programs.

Twenty-eight years later the learning psychologist B. F. Skinner (1954), working from a different psychological tradition, arrived at parallel conclusions and described a machine for teaching elementary arithmetic. Skinner's main emphasis was not upon the machinery for teaching but its psychological basis. He argued that behavior was modified most successfully in the operant conditioning laboratory by arranging for a subject's responses to produce immediate, frequent, positive consequences, while in the classroom, study activities produce few immediate consequences and many errors. In the classroom study of arithmetic, for example, the student does numerous computations but has to wait hours or days to find out how successful he has been. In the laboratory, consequences are made precisely contingent upon the desired responses, while in the classroom large amounts of behavior are emitted without producing any consequences, making it difficult to differentially reinforce the desired behavior. In the laboratory, complex performances are taught by differentially reinforcing responses in the direction of the desired behavior until the final, complete performance is attained, while in the classroom little opportunity is given for gradually approaching complex performances through successive approximations. Skinner had the further insight that arranging positive consequences in the classroom involved the careful and delicate arrangement of subject-matter into sequences designed



to lead students through a series of successive steps to the desired performance. This was the missing link in previous work that has tied current teaching machine endeavor to the problems of teaching specific subject-matter.

In the light of Skinner's analysis, teaching is seen to involve a large number of intimate transactions between the student and the subject-matter, making it impossible for a teacher to oversee the process adequately with more than one student at a time. The role of the teaching machine is to extend the teacher's contact with individuals through teaching programs by mediating this student to subject-matter transaction.

Skinner's original paper on "The science of learning and the art of teaching" (1954) remains the basic document in the field, and his later article on "Teaching machines" (1958) the most erudite and instructive. These papers and others in the field have been collected into a volume edited by Lumsdaine and Glaser (1960). Readers interested in a detailed psychological rationale for teaching machine instruction and reports on the small amount of research that has been done should refer to this volume.

The experimental study of instruction has long followed the experimental-control group paradigm where some groups are taught by the "usual" methods and other groups by one or more experimental variation in content or method. Unfortunately, a great deal of this research has proven to be fruitless. Generalizable results have seldom been obtained because proving that a particular teaching method is superior to one other does not show it is superior to all other methods, and such research has seldom been designed to discover the contribution of underlying variables. However, the greatest failing of

such studies stems from a limited conception of research. In the behavioral sciences, "research" has come to mean the testing of hypotheses and the proving of superiority of treatments. This is usually accomplished in a one-shot experiment that confirms or denies hypotheses and results in entrenchment or reformulation of theoretical positions and the acceptance or rejection of treatments. Although hypothesis testing and "proving" have a legitimate function, they should represent the end of a long line of activities designed to shape up and perfect a treatment or theory. Attempts at premature proof of ideas and operations can very well result in the tragic rejection of valuable contributions. This is especially true of instructional research where the procedures being tested have seldom been adequately perfected, and an experiment serves the dual purposes of diagnosing and proving the worth of the treatment.

Educational researchers should not be blamed for this state of affairs, however, because adequate means of monitoring the instructional process has not been available. Tests and general classroom observations just do not get close enough to the student--subject-matter interchange that is the crux of the learning process. Perhaps the greatest contribution teaching machines can make to education stems from their use as research tools that make a detailed analysis of instruction more feasible and provide a standard method of presentation that can reduce some of the uncontrolled factors introduced by teacher presentation.

The research reported here is concerned primarily with the development of a teaching device and spelling programs rather than with the formal testing of hypotheses or comparative evaluation of the programs. The major role



of the teaching machines has been as research tools to give information about the programs for the sake of further development. Although data on learning outcomes was gathered in the process of investigating other aspects of the teaching machine treatment, a general evaluation of the teaching machine method on the basis of this data would be premature. Future teaching machine research should neither be encouraged nor discouraged by favorable or unfavorable results from a method still in its infancy, but by the discovery of promising variables that will lead to further perfection of the method.

Independent device available not only for research purposes, but also adaptable to everyday school use. The basic operations required of the machine were: presentation of a question, problem, or information; giving a student response; provision for handling the student response in written or some other form; and means of scoring the response. In its simplest form, such a teaching device would consist of questions and answers printed alternately on a sheet of paper with a rack which the student could slip down to expose first a question, which he would answer, then an answer, which he would compare with his own answer for scoring purposes, and so forth down the sheet. Forster and Saper (1958) have used this technique in the teaching of elementary German.

There are several inherent disadvantages in this type of device. With slow students and with some difficult or poorly programmed subject matter, it is impossible to prevent cheating. Answers may be copied rather than being produced by the student, and earlier or later items in a teaching sequence which may "give away" the current item are easily slipped in as accidental or deliberate slips of the student's hand. One of the

## DESIGN OF THE TEACHING DEVICE AND PROGRAMS

### A Simple Teaching Device

Teaching aids and devices have a considerable history beyond the scope of this report. (Interested readers are referred to articles by Mellan (1935), Porter (1957), and the book edited by Lumsdaine and Glaser (1960) for Patent Office references and descriptions of various devices.) The problem faced here was that of providing a rugged and inexpensive device suitable not only for research purposes, but also adaptable to everyday school use. The basic operations required of the machine were: presentation of a question, problem, or information requiring a student response; provisions for handling the student response in written or some other form; and means of scoring the response. In its simplest form, such a teaching device would consist of questions and answers printed alternately on a sheet of paper with a mask which the student would slip down to expose first a question, which he would answer, then an answer, which he would compare with his own answer for scoring purposes, and so forth down the sheet. Ferster and Sapon (1958) have used this technique in the teaching of elementary German.

There are several inherent disadvantages in this sort of device. With some students and with some difficult or poorly programmed subject-matter, it is impossible to prevent cheating. Answers can be copied rather than being produced by the student, and earlier or later items in a teaching sequence which may "give away" the current item are available at an accidental or deliberate slip of the masking sheet. One of the

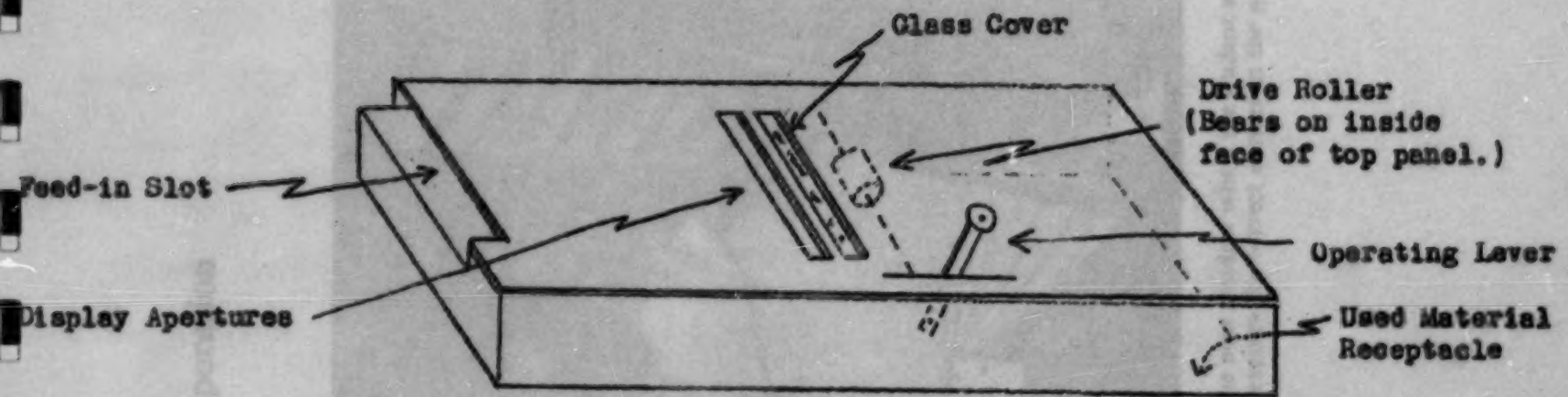


major purposes of a teaching device is to permit controlled, sequential presentation of teaching items with immediate knowledge of results and at the same time prevent premature availability of the correct responses. The basic design to be described below does just that, and little more.

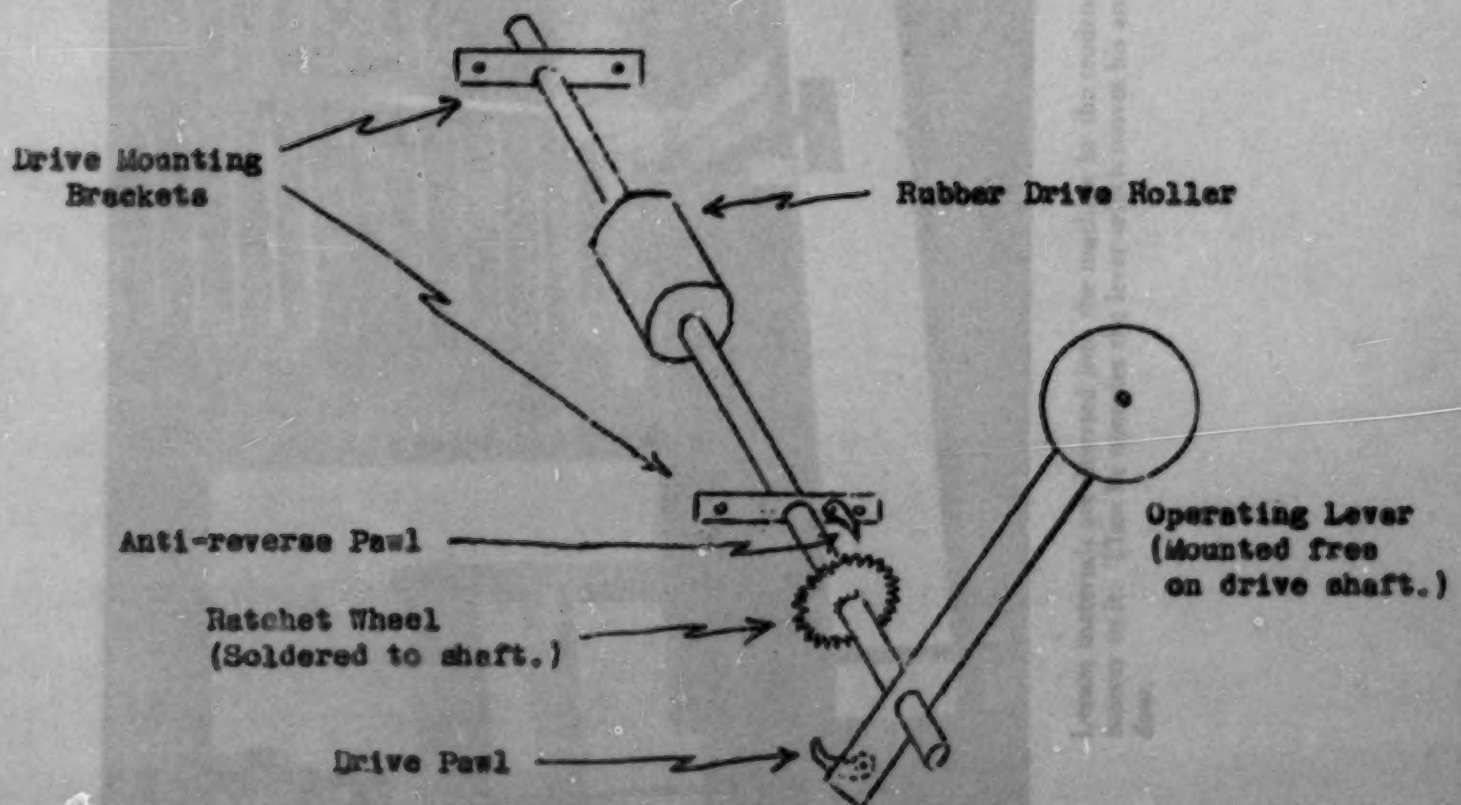
The device is a box about twice the length of the  $8\frac{1}{2}$  x 11 inch paper stock used for the teaching materials, and about 3 inches deep. Materials are fed into the device, one sheet at a time, by the student (refer to FIGURES 1.1 and 2). With a sheet in position, the first teaching item appears in the nearer, uncovered of the two windows. The student makes a response by writing directly on the teaching materials, operates the lever which moves his response under glass and simultaneously exposes the correct answer, compares his answer with the given answer, and scores himself "right" or "wrong." Another operation of the lever brings the next item into position for another cycle. Used materials drop into the far end of the device as each sheet is completed, and are available for analysis by the experimenter or teacher.

Mechanical details of the device are shown in FIGURE 1. The sheets of teaching material are driven through the device and under the windows by means of a uni-directional ratchet drive mechanism, and drop into the far end of the device under compulsion of gravity. The drive mechanism operates in the following manner. (Refer to FIGURE 1.2.) Movements of the operating lever toward the far end of the device engage the spring-loaded drive pawl with the ratchet wheel, which is thereby rotated in a clockwise direction over the travel limit of the operating lever. Since

FIGURE 1: VIEWS OF TEACHING DEVICE



1.1: OVER-ALL VIEW



1.2: DRIVE MECHANISM DETAIL



**Figure 2: The Teaching Device In Operation**



Lesson materials are inserted into the machine by the student. The first item appears in the near window where the student writes an answer to it. Then he operates the lever which moves his answer to the far window and exposes the correct answer in the near window.

the ratchet wheel is soldered to the drive shaft, the shaft also rotates, as does the attached rubber drive roller. The drive roller bears upon the underside of the top panel of the device where it engages the teaching materials. It will be apparent that in this one-roller device satisfactory movement of the paper materials depends upon the drive roller having a greater coefficient of friction than does the panel against which it bears. Any reverse movement of the teaching materials (which would make cheating possible) is prevented by the spring-loaded anti-reverse pawl which is mounted on the frame of the device and triggers against the ratchet wheel. The entire drive mechanism is mounted by four wood screws through the drive mounting brackets and into the wooden frame of the device. Tension adjustments between the drive roller and underside of the top panel are accomplished by means of the hardwood bar mounted transversely on top of the device.

There are several hidden design features which should be described. Transverse clearances for the paper materials are  $1/8$ th inch to forestall any binding within the range of differences encountered in stock  $8\frac{1}{2}$  inch paper. Excessive clearances, on the other hand, would allow materials to feed crookedly and bind. The rubber drive roller is approximately one inch in diameter by one inch in width. A smaller diameter would require excessive operating lever movement for the necessary linear paper travel, and the relatively narrow, centrally mounted drive roller provides a self-centering action as the paper feeds, which would not be available with a full-width roller or a separated pair of rollers. Location of



the drive roller under the material being fed ensures that the tail-end of each sheet will drop into the far end of the device, propelled by the roller, and not remain hung-up to tangle with the next sheet. Clearance between the top end of the paper feed tray and the underside of the top panel is kept at a minimum in order to provide a reading and writing surface as unobstructed as possible. (It will be noted in FIGURE 2 that the student has to write in a slightly unorthodox manner because of the thickness of the top panel. Excessive clearance between the top panel and the paper feed tray would accentuate the problem.) Finally, the paper feed tray is sloped upwards toward the center of the device in order to allow easy feeding in of the sheets and at the same time permit the minimum dimension required for unobstructed writing.

The devices were made in a small wood-metal working shop with no special tools other than temporary jigs for cutting stock. Where metal parts were to be joined, brass stock was used to permit soldering, otherwise, 1/8th inch steel. Panels were 1/8th inch two sided tempered masonite, and frames were 1 x 3 inch pine stock. All wood construction was glued and either nailed or screwed together. Ratchet wheels, pawls, and drive rollers were stock hardware items, but the pawl springs were wound from tempered wire. A crucial part of the manufacture of each device was the testing, running-in, and adjustment required for trouble free operation. Every device required this sort of treatment, and if the devices had been more complex, such treatment would have doubled the cost.

### Redesign

At our current state of knowledge it is not possible to design an optimum teaching device with just the functions necessary for effective teaching and instructional research. Assuming that we knew what functions a device should perform, Rothkopf's (1957) suggestion that the ultimate design of a teaching device be shaped up through a series of successive approximations by simulating device functions in a mock-up situation is probably economical. This would avoid the manufacture of a whole series of machines soon to be outmoded. In order to proceed with research on programming, however, devices of some sort are necessary, and the imperfections of these first approximations to a perfect device can provide valuable hints for redesign. Year long experience with the above device has suggested several changes in design.

A device for teaching program research should have some features not necessary for routine teaching. The major purpose of a research device, in addition to carrying out appropriate teaching operations, is to give the researcher information about the programs he has written, and there are several different sorts of information it is desirable to gather. Of prime importance are the student's responses to the teaching items. These responses, and more particularly the incorrect responses, indicate which portions of the program are inadequate, so that any technique which will allow easy scanning for errors would be a boon to item analysis. In the device used, as in many others, the student scores his own responses by comparing them with the correct responses on the teaching materials and making an identifying mark of some sort, in this case a plus or zero for correct



or incorrect. Incorrect items are picked out by visual scanning for zero marks, but the process could be sped up by arranging for a positive identifying punch in the teaching materials which would permit sorting of responses into correct and incorrect piles by a mechanical device. Such an ideal could be met only at considerable expense and complication, but might very well be economical in the long run if large numbers of responses from many students were to be analyzed. In the present device, item analysis has been considerably simplified by having student responses placed directly on the teaching materials rather than on a parallel paper tape. With this direct response arrangement there is no problem of cross-checking and aligning responses with the appropriate questions during item analysis. For research purposes this arrangement is beneficial, but since the teaching materials are expended, economical reproduction procedures have to be used, and for practical teaching situations unexpended materials with a parallel tape might be preferable. For some research purposes, it is desirable to obtain records of response latency, rate, and total working time. The only attachment necessary for gathering this data is a set of contacts which close each time the device is operated, for connection to suitable recording equipment.

Another important design feature of a teaching device has to do with the preparation, loading, feeding, and security of the teaching materials. Materials used in this research were spirit duplicated on  $8\frac{1}{2}$  x 11 inch paper stock, five teaching frames per sheet. (See APPENDIX A for a sample of the teaching materials.) This form of material has both advantages and

disadvantages. The materials are simple and inexpensive to set up and print on readily available office equipment, but with about 300 separate sheets per program there are problems of collating, handling, and security. Since the materials are inserted into the teaching device one sheet at a time cheating is relatively easy--all a student has to do is to copy the available correct answers onto the teaching materials and then run them into the device--and the sheets can be easily gotten out of order. Although there were few instances of these troubles in a year of experimentation, it is felt that the device should be altered to forestall any such problems. Both the cheating and disorder problems could be solved by having continuous folded or rolled teaching materials which are inserted into the device in lesson-sized hunks. The major difficulty with these formats is that they require expensive and generally unavailable production facilities for continuous printing, folding, or joining of separate sheets. One solution is to retain the present format of separate sheets, but to alter the teaching device so that the materials can be loaded all at once and to provide a mechanism to start the sheets feeding one at a time. Feed mechanisms capable of peeling the top sheet off a pile are generally expensive and complex. The simple friction-feed methods as used in duplicating machines are unreliable at slow speeds and for individual sheets. The simplest expedient is to make use of the student, who is a sophisticated machine, to do the feeding in such a manner that he will not have access to the answers or be able to shuffle the sheets. This can be accomplished by having the collated sheets inserted into the machine in a suitable



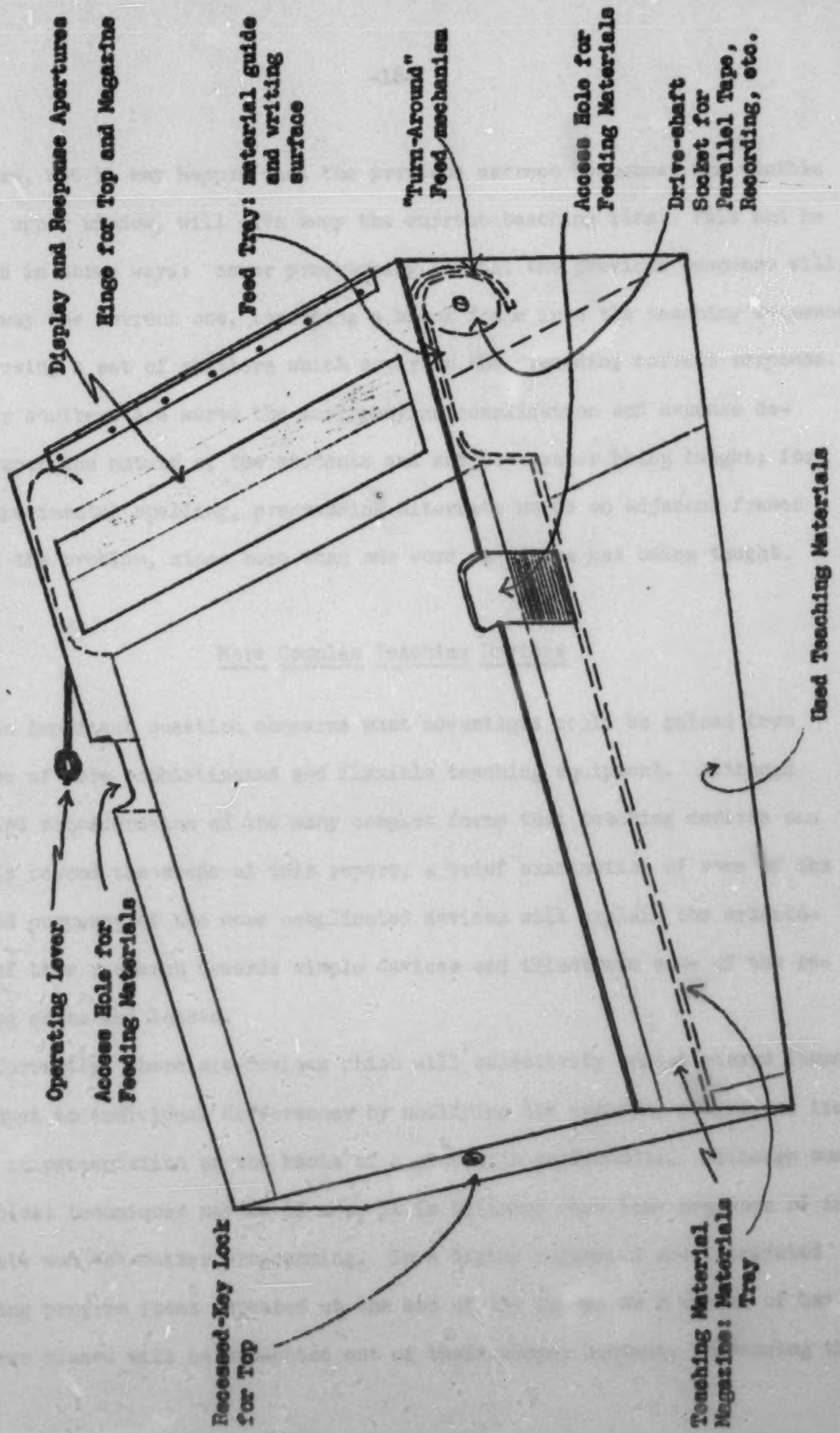
receptable or "magazine," which is then closed. Through access holes on either side of the materials the student can pick off the top sheet with thumbs and forefinger and insert it into the feed-in position (refer to FIGURE 3). It is felt that this technique of handling the teaching materials will preserve the flexibility and economy of separate sheets for research purposes, and at the same time assure ease and security in handling the materials without the expense and troubles of complex feed mechanisms.

Reference to FIGURE 3 will reveal several other suggested modifications of the teaching device. A "turn-around" feeding mechanism, which feeds used materials back under the original stack makes it possible to reduce over-all length of the device by about one-half. (This modification was suggested by Thomas F. Gilbert, who has constructed a pilot model.)

The Drive-Shaft Socket can be used to connect accessory equipment for recording response latency, for triggering a synchronized audio presentation, or for driving a parallel paper tape for receiving student responses with unexpendable teaching material.

One final modification not shown in FIGURE 3 should be considered. As the device stands, a teaching item appears and is answered in the lower of the two Display Apertures; when the machine is operated, the item and its answer move to the upper display window and the correct response appears in the lower aperture ready for comparing with the student's answer. When the comparison and scoring have been made, the machine is operated again, moving the correct response into the upper window and the next teaching item into the lower

**FIGURE 1: REDESIGNED TEACHING DEVICE: X-RAY VIEW**





aperture, but it may happen that the previous correct response, now visible in the upper window, will give away the current teaching item. This can be avoided in three ways: never programming so that the previous response will give away the current one, inserting a blank frame into the teaching sequence, or by using a set of shutters which cover up the preceding correct response. Whether shutters are worth the accompanying complication and expense depends upon the nature of the students and subject-matter being taught; for the experimental spelling, programming alternate words on adjacent frames solved the problem, since more than one word at a time was being taught.

#### More Complex Teaching Devices

An important question concerns what advantages could be gained from the use of more sophisticated and flexible teaching equipment. Although detailed consideration of the many complex forms that teaching devices can take is beyond the scope of this report, a brief examination of some of the alleged purposes of the more complicated devices will explain the orientation of this research towards simple devices and illustrate some of the resulting gains and losses.

Currently, there are devices which will selectively repeat missed items or adjust to individual differences by modifying the teaching program or its order of presentation on the basis of a student's performance. Although such mechanical techniques may be of use, it is believed that they are born of inadequate subject-matter programming. In a highly organized and integrated teaching program items repeated at the end of the lesson as a result of having been missed will be presented out of their proper context, increasing the

possibility of further errors, guessing, or producing the correct response from immediate memory rather than understanding. For a loosely constructed program with little carry-over from item to item (as in rote learning) such a technique may be reasonable and necessary, but not for programs containing tightly worked-out teaching sequences. In such programs, errors should be treated by re-programming rather than by repetition.

Concerning the matter of adjusting a teaching program to individual differences by changes in the content or order of presentation, it can be argued that many observed differences in student performance are a function of differences in learning history which could be erased by highly perfected teaching programs of adequate length and depth. This would make it unnecessary to provide each individual with a unique teaching program. Evidence will also be presented in another section of this report which suggests that the pervasive individual differences often referred to as "mental abilities" show up primarily in the form of differences in rate of performance, a variable adequately handled by individually operated, self-paced teaching machines. In the absence of such perfected programs it is reasonable to use some technique for adjusting to different students in a practical teaching situation, but it also should be realized that such techniques may cover up interesting problems of programming which might eventually obviate the need for individual adjustments in the first place. Consequently, such techniques should be used with caution when engaged in research on programming as opposed to the practical application of the results of programming.

Aside from economic and practical considerations, interest in the nature of spelling as a subject-matter and its programming, rather than the hardware



for presentation, has led to the use of very simple teaching devices in this research. Interest in devices has been limited to deciding what sorts of device characteristics the subject-matter and students require, rather than how the program can be fitted to a particular device.

Despite objections to some forms of equipment, there are certain advantages to be gained from devices more complex than those which have been used here. For spelling, specifically, and probably for certain other subject-matters, there is some point in having devices which do not depend upon a student's judgment for scoring nor upon his handwriting for recording a response.

One basic assumption about programmed materials is that a program builds upon itself--later items in a sequence are successfully answered by a student as a result of previous mastery of earlier items. This assumption of progressive mastery throughout a program may be incompletely met where a student has to evaluate his own responses--as with the device used in this spelling research. It is conceivable that a student using such a device could complete a program never having produced a correct response and never having matched his own responses accurately enough with the given responses to realize the errors committed. Such an event is highly unlikely, but almost every set of response protocols can be expected to produce some mis-scorings, through student "carelessness," "cheating," or other peculiarities of response. It is felt that misscorings should be dealt with in the same manner as individual differences, through re-programming, and by ensuring that a student has been prepared adequately for the program he is requested to study. In spite of all precautions, however, a certain amount of mis-scoring probably will always occur, due to many unpredictable and uncontrol-

lable individual factors. For instance, in the Sixth Grade Spelling Program used with student scored machines, misscorings occurred 1.4 per cent of the time on the average. This is a notably low figure, but the range of misscorings by individual students runs all the way from zero to 25 per cent on particular lessons, indicating the need for some mechanical method of absolutely correct scoring. Devices to receive and objectively score a student's responses via keyboard or other input mechanism are certainly not beyond technical feasibility for subject-matters which require literal textual response, and they should be thoroughly investigated. Not only would they contribute to a solution of the self-scoring problem, but they could also be designed to ensure that a student does not go on to further items in a program until he has answered the current item correctly.

Further advantages would probably be gained if a student were able to insert his spelling responses into the teaching machine with a uniform type face. Figure 4 (p. 22) presents successive samples of the writing of the letter "e" from the protocols of six subjects chosen at random. Each successive instance of the letter for each subject is further along in the section of the program chosen for illustration, and each subject is represented by two samples of writing: one from an early lesson and one from a late lesson in the program. It will be seen that there is considerable instability in the style, size, form, and clarity of the student's writing. There does not seem to be any uniform shift in the quality of the writing from early to late lessons, nor should it be expected, since accurate letter form was not differentially reinforced. Neither is there any consistent change in writing quality within a lesson unless it is towards slightly poorer writing at the end of a lesson. If part of the task of learning to spell



FIGURE 4: SAMPLES OF STUDENT HANDWRITING

| Subject number- | Successive Instances of the Letter "e" |              |
|-----------------|----------------------------------------|--------------|
|                 | Lesson 10                              | Lesson 21    |
| 6e-1            | eeleelee                               | leeklee      |
| 6e-11           | EEElleFV                               | eeleleee     |
| 6e-15           | lelelelele                             | lelelelele   |
|                 | Lesson 13                              | Lesson 28    |
| 6e-2            | lelelelelele                           | lelelelelele |
| 6e-19           | lelelelelele                           | lelelelelele |
| 6e-20           | lelelelelele                           | lelelelelele |

requires accurate visual memory of letters and letter patterns, then much might be gained by providing students with consistent visual stimuli. It is true that the students also need to improve their handwriting, but drawing letters in an acceptable form is a different task from spelling, which involves the sequential selection of correct letters to be drawn. An automatically correcting, multiple choice teaching machine with all letters of the alphabet as alternatives would be ideal for teaching spelling as well as for investigating other aspects of language behavior, and necessary research into the preparation of adequate teaching programs will be hastened by the availability of such devices.

A final mechanical complication to be considered is the addition of an audio stimulus to the teaching machines. Spoken stimuli were not used extensively in this research because of the cost and complication involved, but it is believed that no loss resulted and there is even the possibility of gain. Detailed discussion of the functions of auditory stimuli in spelling is reserved for the section of this report on the nature of spelling as a subject matter, but an outline of the factors involved will be presented here briefly.

The traditional picture of spelling instruction and performance is of an individual sounding out words to himself and taking a dictated test--in general, attempting to translate sounds into appropriate letters. Typical instructions for the study of spelling direct students to "look at the word and say it to yourself, look away and say it again, etc." There are three possible functions of auditory stimuli in such teaching and testing procedures: to inform the student what word he is to spell in the sense of its meaning; to give the student useful phonemic prompts which will help with accurate spelling of the word; and to give a test situation in which spelling performance can be adequately gauged once learning has taken place. It is believed



that the semantic context provided by textual materials in the experimental spelling program performed as well as a vocal context would have, in the case of subjects who could read. Concretely, when a word to be spelled was left blank in a sentence the students knew the word in the sense of being able to attempt its spelling. For certain non-reading students in the second grade a text did not provide adequate semantic prompting, but it is doubtful that such individuals should study spelling. The only potential loss from not having an auditory stimulus available probably rests upon the contribution of sound cues to accurate selection and positioning of letters in the words to be spelled, and this is a complex issue which requires an examination of the letter-sound relationships in English. These relationships will be examined in some detail later in the report.

Had auditory stimuli been available for the spelling program it would have been possible to investigate some of the factors which can only be argued about at this point, and certainly, less text would have been needed to provide an adequate semantic context for prompting student responses. It is recommended that future research into the teaching of spelling make some provision for the use of auditory stimuli, not for practical teaching reasons, but in order to more fully delineate their exact functions.

#### Fragments of a Spelling Program

Three fragments of a spelling curriculum were prepared to investigate several different aspects of teaching machine instruction. Materials were programmed to cover most of the standard spelling curriculum of the second and sixth grades in the experimental school, as well as a small quantity of special work for the fourth grade. The second and sixth grade curricula were taken to be defined by the population of words and exercises in the

standard text series used by the school. The fourth grade materials were especially selected to test the transfer effects of particular programming techniques and so made no attempt to exactly match the school curriculum. (See APPENDIX B for lists of the words taught.) The outside limits of elementary education were chosen rather than adjacent grade levels in order to assess the effects of student age and sophistication upon the techniques of instruction used, and a relatively large portion of the year's curriculum was prepared for both 2nd and 6th grades (34 and 22 out of 34 weeks, respectively).

Such a large quantity of material is seldom prepared for educational experiments, but was here for several reasons: (1) to make it possible to examine some of the results of long-term teaching machine instruction, for instance, the effects upon student performance and motivation of using teaching machines as a normal part of daily classroom procedure throughout the year; (2) to have enough material to produce an effect of practical magnitude, comparable to that usually provided by a normal year of instruction; (3) because preliminary estimates indicated that the self-teaching materials would be used at a rapid rate, it was necessary to prepare sufficient material to ensure adequate contact with the student's total in-school study time; and, (4) because the research was concerned primarily with the nature of the programs generated, it was necessary to provide sufficient materials to make comparisons between parts of the programs. It was the commitment to experiment with a large part of the spelling curriculum that led to the programming of existing materials rather than the generation of completely new lessons. Although the adoption of an established spelling series placed certain undesired constraints on the program, in the selection of words and their presentation,

other word structure exercises, the definitional items, and in some cases, further sentence contexts, were written explicitly for the experiment.



it was felt that this was offset by the advantages of having a large quantity of material and by a responsibility to protect the rights of student subjects who would be expected to fit into the normal sequence of the school's spelling curriculum at the close of the experiment.

Although there are many variations from lesson to lesson in the exact programming techniques which were used, the basic pattern of programming for the 2nd and 6th grade materials is described below (see APPENDIX A for a sample lesson of the programmed materials). First of all, the student identifies the spelling words by underlining them in a meaningful sentence context consisting of a story to be read. After having responded to a series of such items, the student answers a question on the content of the story. This is to ensure that he reads the story in addition to underlining the appropriate words. Third, the student is required to make some discriminations about the letter structure and/or sounds of the spelling words; for instance, picking out rhymes, doubled letters, or tricky letter combinations such as "ei" rather than "ie." Then he matches the appropriate spelling words to given dictionary definitions in order to ensure understanding of the words to be spelled. Finally, after correct meanings of the words have been reinforced and peculiarities of their structure have been responded to, the initial sentence contexts are again used to evoke the spelling words, printed with missing letters, which are to be filled in by the students. Further frames of this same nature require different letters to be replaced, more complete responses, or review, depending upon the particular words being studied. In all of this programming, the sentence contexts were drawn directly from the regular school spelling texts, as were a number of the exercises calling attention to word structure. Other word structure exercises, the definitional items, and in some cases, further sentence contexts, were written explicitly for the experiment.

Letters omitted from the words to be spelled were determined by a number of factors. Generally, all of the letters of a word were seldom omitted, rather, primary emphasis was placed upon parts of the words it was assumed would be misspelled because of poor correspondence between the letters and sound cues available from spoken English. Actual writing of the items was guided by data on the regularity of English spelling, obtained from an analysis by Moore (1951) and data on the most frequent misspellings of words, supplied by Gates (1937). However, it was found that the writer, as a native speaker of English, usually could make adequate judgments about what parts of the words to emphasize, and the Moore and Gates data were used primarily as checks. The sequence of letter omissions from item to item for spelling a particular word usually started with "easy" letters--those with regular sound cues--and ended with letters which could not be accurately predicted on the basis of sound alone. (See the sequence on "steady," APPENDIX A, sheets 6.31-8 and 9). This sort of pattern gave students an opportunity to observe the more difficult parts of the words while spelling the easy parts, and it was hypothesized that such observations, if they took place, would increase the probability of spelling the words correctly.

The sequence of item presentation was constrained by the teaching machines, which had no shutters. This meant that the correct answer to the just previous item was always available so that if two or more items in a sequence were concerned with the same word, its spelling would be given away. This problem was handled by dealing with spelling words in five word sequences in which no two items in a row were concerned with the same word. The choice of five word sequences was accidental in the sense that only five question and answer frames could be physically accommodated by one sheet of the programmed material, but there probably were positive consequences from this arrangement. By presenting



relatively long, related sequences of semantic context the probability of trying to spell the appropriate word was probably increased, and separation of occurrences of the same word probably tended to break up any immediate rote memory effects which would have been produced by numerous repetitions of the same spelling word in a short time interval.

Although the 4th grade program bears a close relationship to the 2nd and 6th grade programs, it differs in many respects. It contains only 456 items as compared to 1505 items in the 6th grade program, and 1015 in the 2nd. A comparison of items per word taught, however, reveals that the 4th grade program uses 5.7 items per word as compared with 4.6 and 4.0 for the 2nd and 6th grade programs, respectively, an increase of 24 per cent and 43 per cent in the amount of teaching material per word. The 4th grade program differs also in how the words to be taught were selected. Rather than using the established curriculum, words were chosen to: (1) be a part of the students' understanding vocabulary; (2) not be a part of the spelling curriculum already studied; and (3) prevent certain specific spelling difficulties. Although average grade placement of the words chosen is 4.9, the range is 3.0 to 8.0, and 14 per cent of the words were chosen at or below the 4th grade level, from a list of words already taught in the normal curriculum, in order to provide a "buffer" set of relatively easy materials at the beginning of the program. Other 4th grade level materials, which had not been previously taught, were distributed throughout the program to prevent any concentration of difficult items.

The basic pattern of programming in the 4th grade also differs from that used in the 2nd and 6th (see APPENDIX A for a sample of the materials). The first 45 items are "alphabetical review," consisting of letters to be placed into the order of the alphabet and alphabetical sequences with missing letters to be replaced. This is followed by the main body of the program, consisting of

words to be spelled, arranged so that each sheet of programmed material takes up one word. In the later half of the program 34 of the 70 words taught are introduced a second time on "review" sheets which take up three words each. In the case of sheets of programmed material dealing with one word, the problem of two or more items in sequence "giving away the answer" was usually handled by having adjacent items consider different aspects of the word to be spelled; e.g., one presentation of "direction" required students to replace missing letters in "d\_r\_ction," while the next required "direc\_ \_ \_" to be filled in.

Two versions of the 4th grade program were prepared, differing in the exact type of teaching item used for certain words. Both versions begin with the alphabetical review, continue with a common "buffer" set of 10 words presented in the same manner, and both versions present the same review items. Beginning with item 97, however, 35 of the words are treated differently in the two programs. Each word to be spelled is introduced, underlined, in a sentence context; in the control version, students write missing letters into the word or respond with the number of letters in the word; in the experimental version, students make a judgement about which of two groups of words the word to be spelled looks most like or respond with letters the word to be spelled has in common with another group of words. One or two items of this sort are included in the series of items for each word treated differentially in the two versions of the program. Although this amounts to only 41 out of a total of 169 items, many are concerned with the same families of spelling words, for instance, those having "-ou-" in them, so it was hypothesized that there would be a significant cumulative effect over the program. APPENDIX B contains a list of all the words taught in the 4th grade program, and those which were differentially treated are marked with a plus (+). APPENDIX A contains sample sheets



from the program showing different treatments of the word "should" in the experimental and control versions of the program along with a sheet of alphabetical review and spelling review from the later part of the program.

### Revision

The teaching programs prepared for this research, as well as the teaching machines, represent only a first approximation to suitable, operational materials. It was expected that they would have to be revised before further use, and ample evidence has been gathered to support this notion. One of the benefits of self-instruction programs is that they can be revised on the basis of concrete information, provided by many explicit student responses. Indeed, the producer of such materials probably has a moral obligation to engage in revisions if he intends to release the materials for practical teaching use. Although it is not planned to carry out further work with the present set of materials, an examination of the revisions which could be made will be instructive to others planning to write spelling programs.

There are at least four bases for the revision of teaching programs. Pretest-posttest data can indicate whether a program teaches what it aims to teach, or not; so long as expected and adequate gains are registered and there are no undesirable side-effects, it can be assumed that a program does not require revision. There are two possible exceptions: if it is deemed that the wrong thing is being taught, or if comparison with some outside, independent standard indicates that efficiency of the program can be materially increased, then revision is in order. Once it has been decided to revise, the actual revision operations can be based upon data provided by the pre- and posttests, upon student responses to items in the program itself, and upon the program

writer's conception about what particular programming techniques are appropriate. Certain characteristics of each of these bases for revision require detailed examination.

Although a test which systematically samples the contents of a teaching program may be a suitable indicator of the general success of the program and may be able to give indications of what areas, in general, need revision, it cannot provide detailed information about how individual teaching items should be revised. The obvious source of this information is student response protocols from the original teaching materials and the obvious procedure is to look for items which produce errors, make inferences about why the errors were made, and revise. Though this sort of procedure may eliminate errors in the responses to most teaching items, it will not guarantee a program which produces the desired student learning. In teaching programs carefully designed to lead a student, step by step, from ignorance to competence, all items are not of equal importance. Certain items, call them criterion items, most closely require the terminal behavior it is desired to teach, and, generally, such items come near the end of a sequence of items designed to prepare for and lead into the final performance. In the spelling programs, such items are represented by the semantic context, structural, and definitional items which lead, finally, into items which require direct spelling responses. The important factor for program revision is that the criterion items often are answered in an extended context provided by previous items, and that although these antecedent items may be a necessary part of the teaching procedure they may not be a part of the environment which will naturally occur as discriminative stimuli at a time when the taught performance is appropriate. Even if a program is so sophisticated that no formal, "unnatural," response prompt-



ing stimuli are present in the criterion items, and terminal behavior is emitted only in the presence of stimuli which are known to occur under "real life" performance circumstances, it is still likely that, by virtue of temporal contiguity, responses will be partly under the control of stimuli which have been presented in earlier parts of the teaching sequence. Programming techniques may overcome this problem by breaking items loose from their original contexts and salting them into sections of a program where they have no direct relevance, or by requiring application of the terminal behavior to new problems or situations. Here, however, we are concerned with how one can evaluate whether such techniques have been adequately applied to a program. The gist of the argument is that revision procedures must determine, in some way, whether correct student responses to criterion items really mean that terminal behavior has come under appropriate stimulus control, and that responses made to criterion items within the context of a program will not give this information. Since a test generally contains items designed to evoke responses by means of minimal within and across item cues, criterion items selected directly from a program and presented in test form should provide the type of information sought here. In such a test, the items would appear out of their original teaching context, and since students would not have just previously responded to the lead-in items, minimal response support would be available from irrelevant stimuli. This suggests that the most effective sort of empirical program revision data would be provided by a joint analysis of student responses to teaching items within a program and to testing items within a test made up of criterion items taken from a program. Responses to items presented within the program context would give information on technical excellence of the programming, while the same items in test form would provide data on what might be called "stimulus

control validity" of the program. An alternative source of criterion items for such a test in the universe from which criterion items in the program were originally drawn. However, a test made up of such items would be a test of "transfer," and would be inappropriate for program revision unless there was reason to believe that common factors in the universe of items would lead to correct behavior toward them all after exposure to only some.

A final basis for program revision hardly needs to be mentioned: the programmer's conception as to what techniques are appropriate. In the above discussion, it has been assumed that a program should be written to eliminate student response errors. Although this is a perfectly reasonable assumption to make about some programs, it is conceivable that errors would be unavoidable or even desirable in some cases. For instance, in designing a program for retraining individuals with varying histories it might be more efficient to bring out wrong responses and extinguish them than to attempt a complex series of items to bring out the correct responses without error. Or, if students have to learn a complex performance which will take place under conditions where they are likely to make errors, it may be reasonable to simulate these conditions in the teaching materials in order to teach "frustration tolerance," "perseveration," and how to obtain information from negative instances. If a program is prepared with these considerations in mind, it is reasonable to look for "errors," rather than correct responses, in an item analysis for revision purposes.

Many of the above considerations will be relevant when experimental results from the spelling programs are examined and specific suggestions are made for revision.



## STUDIES IN THE ANALYSIS AND TEACHING OF SPELLING

Four separate but related studies were performed with the teaching machines and programs. The 2nd, 4th, and 6th grade programs were administered to classes in the experimental school, where control data was also gathered. Experimentation covered the entire 1957-58 school year. In addition, an analysis was made of the sound and letter structure of a corpus of English words from which items for the programs were drawn. A detailed description of the problems studied, procedures used, and results obtained follows below.

### The Problems Studied

Aside from testing operational aspects of the teaching machines, which has already been discussed at some length, the long period of experimentation allowed investigation of several salient features of the teaching machine treatment. In fact, some of the problems investigated did not appear until well into the experiment, but the variety of classes and materials involved permitted considerable leeway for taking these problems up, and despite the fact that reasonable controls were exerted, the main feature of the experiments was their exploratory nature. The term "exploratory" is used here to denote research in which the major emphasis was placed upon the discovery and development of new ideas and techniques rather than the proving and demonstration of old ones. Although some conventional measurement techniques are used, the pattern of the research follows more closely that described by Skinner (1956) than the conventional paradigms of proof.

Teaching efficiency.—Although teaching efficiency is the first and foremost problem investigated in most educational research, it was not the primary interest here. This research was concerned with the development of a new technique, still far from being perfected. However, in the process of investigating the other problems to be discussed below, data were gathered on the efficiency of the teaching machine treatment compared to one other, more conventional, teaching technique, and the results will be presented. The comparative data also served as an indicator of whether students were being deprived of their rights to learn.

Teaching efficiency, in the sense that it can be translated into dollar statements, is vitally important to school administrators. Ultimately, it concerns such diverse things as student study time per unit of learning, quality of learning outcomes, teacher time, material cost, and many other factors, but it is not profitable to study this complex of factors until a perfected teaching instrument is available. Here, the study of efficiency has been limited to a measurement of spelling achievement relative to typical achievement norms and norms provided by control groups, as well as measures of study time required. Data were provided by the annual achievement tests administered by the school system at the end of each year as well as by a different set of achievement tests chosen especially for this experiment. In addition, there were weekly tests and periodic tests which specifically sampled the spelling words taught.

On the basis of preliminary studies it was hypothesized that the experimental treatment would:

- 1) Produce greater spelling achievement than that produced by the control treatment.



- 2) Result in a year's spelling achievement greater than that expected according to achievement test norms, which are set up to record a grade equivalent gain of one point for each year in school.
- 3) Produce a narrower range of achievement than that produced by the control treatment.
- 4) Teach the spelling materials in a shorter time than the control treatment.

There were several sources for these hypotheses. Previous studies of teaching which manipulated knowledge of results (KOR) have generally reported superior gains for the experimental groups which received immediate KOR (Ammons, 1954; Porter, 1957). The same effect was expected here, but no attempt was made to get at the complex of possible causes for superior achievement, such as immediacy and type of KOR, the effects of having to make an explicit response, etc. The predicted time saving in the teaching machine treatment was based upon the observation that students (Ss) in a tutorial situation generally work faster and more continuously than in solitary study, and one crucial factor in tutoring is that the student (S) receives immediate KOR from the tutor. A similar prediction can be generated from work done on schedules of reinforcement, where it has been observed that intermittent reinforcement often generates high rates of responding (Farster and Skinner, 1959). In the teaching programs, KOR is placed upon an intermittent schedule by virtue of the varying lengths of the teaching items, and the varying times it takes Ss to complete them. It was thought that a narrower range of achievement would be produced by the teaching machine programs because the careful, sequential structure of the programs would raise the achievement level of Ss who had tended to have difficulty with spelling.

Year-long changes in performance.--In any experiment extending over a considerable period of time one can expect to observe certain fluctuations and, perhaps, systematic changes in the performance of subjects. Often, such changes are ignored by gathering only pretest and posttest data or are washed out by averaging individual performance records. When prominent irregularities do appear they are deplored for the difficulties they introduce into data analysis and interpretation, for if the experimental treatment has been kept constant and other causes cannot be identified, nothing is left but to attribute the changes to motivational effects, e.g., the subjects became "tired," "lost interest," or worked beyond their "attention span." Such explanations, however, do not help to increase control over experiments and teaching for they do not point out how to keep subjects interested or how to extend their attention span. Since the main objective of this research was to discover controlling variables in the teaching machine treatment and to investigate how they operated, careful attention was paid to long-term changes in performance and its correlates.

It was expected that long-term changes in performance would reflect three major factors: the effects of merely engaging in an experiment, the effects of working week after week on teaching materials constructed on the same general pattern, and the results of specific operations carried out by the teacher and experimenter.

Most educational experiments are plagued by the so-called "Hawthorne Effect," the motivational results of merely engaging in an experiment irrespective of the differential treatments carried out, and elaborate procedures are sometimes devised to control the effect. It has been suggested by many persons that teaching machine instruction would provide an especially strong case



of the effect because of the "novelty" of the teaching technique used, but that with continued use the novelty would "wear off," and learning gains would decrease. This conjecture is supported by experiments on exploratory behavior, where it has been shown that infra-human primates will perform various complex tasks in order to expose themselves to novel stimuli, but that the effect decreases with increased contact with the stimuli (Harlow, 1950; Berlyne, 1960). It was expected that if such factors were operating in the spelling treatment, achievement would be high at first, followed by decreasing achievement throughout the year. Also analogous to the animal situation, it was thought that any activities such as vacations, regular teaching, no spelling for a week, etc., which intervened between sections of the spelling program routine would result in a temporary increase in achievement (cf. Harlow, 1950).

It has been hypothesized by many researchers who are interested in teaching machines as a generally useful device for educational research, that such instruments should enable greater control of instructional experiments by eliminating the variability introduced by teachers. It is probably too much to expect, however, that any teaching device in daily classroom use, administered by a teacher, would be immune to some measure of teacher control. Indeed, investigations of operant behavior in school-age children, analogous to teaching machine work in that the Ss interact with instruments rather than teachers, have found that social interaction with the experimenter remains a potent controlling factor that has to be arranged carefully if valid data are to be gathered (Bijou, 1957). Insofar as the teaching machines were used in a situation which allowed social intercourse between students, further, relatively uncontrolled, effects upon long-term consistency of student per-

formance were expected. It was realized that adequate control of such factors could not be obtained without disrupting the normal course of classroom organization, so anecdotal records were kept of all major perturbations of student performance observed as well as the associated circumstances. In contrast, at several points the teachers gave directions deliberately designed to influence student behavior toward the teaching machine lessons, and the effects of these operations were also noted. Week to week fluctuations in student spelling performance were expected to reflect the operation of such personal factors in the classroom and, in cases where antecedent conditions could be identified, give some indication of how susceptible teaching machine instruction is to the depredations of such uncontrolled factors.

The prime data used for tapping changes in performance were provided by weekly spelling tests, and the effects of novel teaching techniques, participation in an experiment, and teacher instructions were sought in changes of weekly achievement. Secondary data was provided by a standardized achievement test given midway through the experiment, but this test was expected to be insensitive to the more transitory changes in performance brought about by weekly happenings in the classroom.

Individual variables.--Perhaps the safest prediction one can make about the outcome of a teaching treatment is that some individuals will profit more than others, and a long tradition in educational psychology has been concerned with predicting who should receive what treatment. Little practical progress has been made, however, in providing special treatments for different categories of students, except for the extreme cases of physical and mental deviation, where individual instruction is given. There are at least two



reasons for this state of affairs: the predictive instruments used have not been specific enough to the subject-matter being taught to give proper diagnostic information, and when the information has been forthcoming it has indicated treatments so specific that teaching programs have had to be prepared for almost each individual, a gigantic undertaking. In this situation, one of the generally held hypotheses about teaching machine instruction is extremely attractive: that it can better accommodate individual differences than standard teaching techniques. Insofar as teaching machine instruction utilizes techniques such as self-pacing and strongly prompted items to maximize the possibility of correct student responses, and correct responses mean that learning is taking place, this seems to be a reasonable hypothesis.

Considering the nature of teaching machine programs, the most likely manner of accommodating a wider than typical range of individual differences in ability is through success with the lower end of the ability continuum; no matter how bad a teaching technique is, the better students seem to learn if the critical material is presented. Of course, a teaching program can be extended in the direction of more and more sophisticated subject-matter content, and so be useful to individuals with superior ability who have already mastered the more elementary aspects of a subject, or a program can be extended downwards to reach students who are unsophisticated by virtue of limited experience or little profit from their experiences, but this is not the kind of accommodation of individual differences meant here. Given two teaching programs which attempt to teach the same terminal behavior, it is possible for one of the programs to be more successful with a wide range of individuals because the propaedeutical items and operations are more suitably arranged. Since the experimental and control programs used in this research

covered the same terminal behavior, any advantages in accommodating individuals were expected to accrue from such programming factors, rather than from subject-matter coverage.

One possible consequence of using a teaching program which accommodates a wide range of individual differences is that the correlations between learning outcomes and typical predictors of achievement such as IQ scores and past achievement will be altered. If individuals on the lower half of the IQ score continuum achieve more than is typically expected, as has been hypothesized, one would expect to see the correlation between IQ scores and achievement lowered. This sort of result can be further rationalized by recognizing that IQ tests are designed to predict achievement with the teaching techniques usually used in school, which require Ss to persevere in verbal tasks with very infrequent KOR for reinforcement. Any test which does not give KOR, and tests are typically designed not to give KOR, implicitly measures performance under non-reinforcement conditions, and such conditions are not relevant to teaching machine instruction.

Another possible consequence of teaching machine instruction is that it may increase the predictive power of certain individual variables which normally bear little relationship to achievement. For example, in the spelling programs used, Ss obtain KOR by careful matching of their written responses to the correct answers provided, so it might be expected that measures of handwriting quality and careful matching behavior would be good predictors of achievement.

The performance of a number of individual students were examined in order to tease out some of the many other possible causes for individual differences



in achievement. Age of the students, whether they typically cheated or not, the characteristic number of errors made on the program, misjudging correctness of responses, and liking for the mode of instruction were all thought to bear possible relationship to achievement, and were investigated for both individuals and the groups.

Programming factors.—Three major kinds of programming problems were studied in this research. One problem was the nature of spelling and how it affected the programming techniques used and learning outcomes obtained; this is discussed in a section of the report to follow. Another problem was the basis for analysis and revision of a program after student tryout; this has already been discussed on pages 30-33 of the report. The problem to be discussed here concerns the relationship between certain general characteristics of the spelling programs and student responses to the programs.

Since programming of the sort reported here is still in its infancy, no well validated and agreed upon rules have been set up to guide the design of such teaching materials. However, among various groups of workers in the field, several different sets of assumptions and working principles have been developed, and generally are followed in the construction of practical teaching programs. The principles followed here are representative of those used by individuals who claim to follow the notions described by Skinner (1958) and others (cf. Holland, 1960). Stated quite simply, the programming principles followed are: (1) provision of immediate reinforcement through knowledge of results; (2) leading Ss gradually to a complex performance through a series of successive approximations; (3) providing sufficient stimulus control for each item in a teaching sequence to maximize

the probability of a correct response; and (4) ensuring that stimulus control initially used to "get out" correct responses has been transferred to stimuli which will be found in the environment when the taught performance is appropriate. (Other principles not directly relevant to the spelling program have been omitted.)

With materials designed according to the foregoing specifications it is assumed that students learn by correctly responding to individual items under control of appropriate discriminative stimuli provided by the teaching sequence used. It follows that errors or forms of response that alter the desired stimulus control should have deleterious effect upon learning outcomes, and such effects can be investigated by looking for appropriate correlations. If the programming assumptions are correct, sections of a program which produce a high frequency of student errors should result in lower learning outcomes than sections with low errors, and the same should be true for individual students, those who make many errors should learn less than those who make few errors. A student who "cheats" the teaching program may do so by any number of techniques which allow him to copy the correct answers. Essentially, cheating breaks the stimulus control built into a program by enabling a student to respond under the control of inappropriate cues. In the case of spelling, the terminal behavior one attempts to set up consists of writing the correct letters of a word under control of sound, structural, and other cues, and this process is short-circuited by the opportunity to copy. Other factors being held constant, one would expect individuals who cheat to learn less than those who do not because their performance will not have come under control of appropriate discriminative stimuli.



In the teaching procedures used, Ss scored their own responses by comparing what they had written with the correct answers provided by the program. Although a great majority of the comparisons were made correctly, some individuals and sections of the programs produced more miscores than others. Several different forms of miscoreing are possible and, perhaps, attributable to different factors. In the case of wrong responses consistently scored as correct, one can infer that such behavior is not accidental but probably under control of the aversive consequences of being recognized by the teacher or peers as one who makes mistakes. Individuals who made this sort of miscore were probably highly motivated, but the miscores were generated by sections of the program which were poorly done, in the sense that they created errors, and should produce the same effect upon learning outcomes as errors scored correctly. Indiscriminate scoring, calling correct and incorrect responses right and wrong with equal frequency, corresponds to what is usually called "carelessness." Calling such behavior careless does not explain it, however, and its origins should be sought in the palpable variables of the experiment. One possible explanation, which can be sought in the data gathered, is that careless observing behavior is generated by items which can be responded to correctly most of the time without careful observation. Under such conditions, careful observing behavior would be expected to extinguish and other behavior be shaped up instead. Such things as looking away from the task at hand, day-dreaming, and so forth are likely to increase in frequency under such circumstances and further increase the chance of scoring errors. It was hypothesized that this sort of careless behavior should be found most often in the records of students who already know most of the

spelling material presented and to whom the task would seem "boring" (cf. Holland, 1958) and least often in the records of students who did not know the material but were able to use the program without many errors. A rather extreme form of misscoring is represented by an individual who marks correct responses as though they were incorrect. Consistent behavior of this sort, as contrasted with the case of marking everything incorrect, requires accurate observation of the correct matches between the written and supplied responses. The only feasible source of such behavior would be some sort of extra-program contingency such as special attention from a teacher who provided "extra help" when errors were made. It is entirely possible for such treatment to set up a reinforcing contingency which would maintain undesirable behavior (Henry, 1957).

A final programming variable considered was the effect of the amount of teaching per spelling word upon learning outcomes. Keeping the difficulty of the teaching materials constant, there are several possible relationships between the number of teaching items per word and the amount learned. Studies of rote learning and the acquisition of skills typically produce a positive, decelerated relationship, the more or less standard learning curve. It was not expected that such a relationship would be generated by a well written teaching machine program. Each trial in rote memorization or skill acquisition requires production of a response made under minimal support from the environment, and cues to the correct form of response in early trials are provided solely by means of the confirmation supplied after a response has been made. As a consequence, a repertoire of incorrect responses is developed which must extinguish while the correct repertoire is built up, a long process



which results in the typical learning curve. Teaching machine programs are designed explicitly to avoid such a condition by providing the necessary stimulus support in early trials and then removing it or substituting other stimuli for it in such a way that subjects are still able to make the correct responses under performance circumstances. Under this procedure, the number of items in a teaching sequence is determined by what steps the programmer feels are necessary to lead a student to the final, correct performance. One kind of behavior may take twice as many items to teach as another, but there should be no difference in the ultimate performance produced. Hypothetically, it is even possible that the shorter of two sequences designed to teach the same thing would produce superior results. In the experimental spelling programs it was anticipated that there would be no relationship between the number of items per word and learning outcomes. Insofar as a positive relationship does exist, it is probably a measure of the programmer's lack of success in producing anything more than a rote program.

On English spelling.--Unfortunately, most research on the teaching of spelling has suffered from the same ills that have plagued other instructional research: concentration upon gross "teaching methods" rather than upon the underlying psychological variables, and inadequate consideration of the nature and organization of the subject-matter being taught. In spelling, for instance, much attention has been paid to such matters as how many tests a week to give, whether to present words in lists or sentences, etc., and although considerable work has been done on specifying what words should be taught at what level, little actual analysis of the sound-letter relationships of English has been undertaken (Horn, 1960, pp. 1337-1354).

A systematic review of research on spelling will not be presented here, for others have more than adequately done the job (Horn, 1960; Hanna & Hanna, 1959). Instead, the nature of English spelling will be examined and the character of spelling performance analyzed for the sake of generating hypotheses about how to effectively program spelling instruction.

The major difficulty in learning to spell English lies in the complex and often unpredictable relationships between the forty or so phonemes of English and their letter representations, generally called "phonics" by teachers of reading. The nature of these relationships makes it impossible, in many cases, to predict the spelling of a word on the basis of its sound alone, e.g., from phonics considerations, /fortiyn/, correctly spelled "fourteen," can be represented by all of the letters shown in TABLE 1. The figures beside alternative spellings are based upon the percentage of occurrences of the spellings and indicate the probability of being correct when guessing upon the basis of individual sound cues alone. A reasonable spelling strategy to adopt under such circumstances is to guess the most frequent spelling, which can often be done by a literate person. A student who guesses that /f/ is spelled "f" will be correct most of the time, but there is no clear cut advantage to guessing that /iy/ is spelled "ea," because alternative spellings are almost equally probable. Spelled in this manner, /fortiyn/ comes out to be "fortean," reasonable, but incorrect.

The most direct method for examining the causes of spelling difficulty in English is to analyze the sound-letter relationships involved for the number of alternative spellings of the same sound and the frequency with which they occur. Moore's Dissertation (1951), reported in Hanna and Moore (1953),



TABLE 1

VARIANT PHONIC SPELLINGS OF THE WORD "FOURTEEN"  
AND PROBABILITY OF THE SPELLINGS BEING CORRECT

| Phoneme | /t/      |             | /o/      |             | /r/      |             | /ɹ/      |             | /ty/     |             | /n/      |             |
|---------|----------|-------------|----------|-------------|----------|-------------|----------|-------------|----------|-------------|----------|-------------|
|         | spelling | probability | spelling | probability | spelling | probability | spelling | probability | spelling | probability | spelling | probability |
|         | r        | .84*        | o        | .48         | r        | 1.00        | t        | .82         | oa       | .31         | n        | .93         |
|         | ff       | .08         | o-o      | .26         |          |             | te       | .10         | ee       | .26         | na       | .04         |
|         | ph       | .05         | ow       | .14         |          |             | od       | .04         | e        | .20         |          |             |
|         | gh       | #           | oa       | .08         |          |             | tt       | .03         | o-e      | .07         |          |             |
|         |          |             | ou       | #           |          |             |          | #           |          | #           |          | #           |
| Totals  |          | .97+        |          | .96+        |          | 1.00        |          | .99         |          | .84         |          | .97         |

Data adapted from Horn, 1957; rounded off to two significant figures.

#Very low figure. Not listed by Horn

and Horn's paper (1957) present relevant data. Hanna and Moore, citing results from Moore's dissertation, state that 75 percent of the vowel phonemes are spelled by their "regular" (quotation marks mine) letter representations from 57-99 percent of the times they occur. This does not seem to be highly irregular spelling. However, because of some peculiarities of phonemicization, use of a limited corpus of words (3000 word basic spelling vocabulary), and methods that did not count doubled and silent letters as irregularities of spelling, Moore's data gives a rosier picture of the regularity of English than is the case. Reducing the 26 "vowel speech sound" categories in

Moore's dissertation to 20 to conform with common linguistic practice, seven is the median number of different spellings in each, with a range of 2 to 15. Eight of the 20 categories, or 40 percent instead of 75 percent, are now spelled by their "regular" letter representations 57 percent of the time or more. This implies considerably less regularity in spelling than the interpretation supplied by Hanna and Moore. In addition, the concept of "regularity" needs further examination. As Horn has pointed out, Hanna and Moore use a special definition of "regular" spelling--the most frequent spelling. In TABLE 1, for instance, the "regular" spelling of /iy/ is "ea," which occurs with a probability of only .05 greater than the next spelling, "ee." It is doubtful whether this difference is of any practical value. In a society where misspelling is unacceptable, and a sound has more than one possible spelling, guessing will not pay off all the time. Depending upon the relative frequencies of occurrence of the alternative spellings, guessing will lead to more or less spelling errors. Whether the most frequent spelling of a sound is to be called the "regular" spelling or not depends upon establishing some acceptable level of errors in spelling performance. If five percent error is acceptable, then the "regular" spelling has to be 20 times as likely to occur as the alternative spellings possible. Only two of Moore's most frequent ("regular") vowel sound spellings meet this criterion, and it is probably not a rigorous enough criterion for our present social standards.

This detailed analysis of Moore's dissertation data has been carried out to give some indication of the difficulty of the spelling task in English. Indeed, it seems remarkable that people spell as well as they do in



the face of such difficulties. Actually, the picture has been painted darker than it is. The spelling of consonant sounds is considerably more systematic than the spelling of the vowels; there are bases for spelling other than direct sound-letter translation; and there are higher-order regularities in English spelling, some of them coded in the form of "spelling rules," but many responded to by good spellers without verbal mediation.

Adequate assessment of the higher order regularities in English spelling is a gigantic undertaking which involves not only the sorting out of individual sound-letter relationships but also any other contingencies which are correlated with these relationships. The position in the syllable or word of a sound to be spelled often correlates with one or more of the alternative spellings, and semantic considerations bear a close relationship to many spellings, as in certain roots and affixes. For example, the following second order contingencies, if known by a student, would tend to increase his chances of correctly spelling /fortiyn/. Since no English words begin with the "ff" or "gh" spellings of /f/, chances are about 84/5 or 17/1 that the spelling is "f" rather than the alternative, "ph" (see TABLE 1). Without consideration of the positional contingency, chances of being correct drop to about 6/1. Several factors operate to determine the spelling of /iy/. If the meaning of /fortiyn/ is known to the speller, as it should be when he is attempting to use the word, the syllable /tiyn/ immediately falls into a class of words with common reference and spelling. If correct spelling of other members of the class is known, there is a good basis for guessing the spelling of /fortiyn/, even if it has never been attempted before.

Perhaps the greatest contribution to correct spelling in the face of ir-

regular phonics is made by the contingencies in sequences of sounds and letters. That is, in the spelling of some words, the sounds (and letters) preceding and following certain sounds predict the spelling of the sounds. In order to discover the sequential contingencies useful in spelling of a sound, for example, /iy/ as in "leak," it is necessary to tally the occurrences of "ea" versus "ee," etc., after /l/ and before /k/, etc., in a representative sample of words. Based upon a count of all the monosyllables in Thorndike's first 6000 words (1944), probabilities of the various spellings of /iy/ are given in TABLE 2. Contingencies that produced no useful differential frequencies have been omitted. It is clear that the chance of guessing the correct spelling of /iy/ in "leak" would be increased by knowledge of these contingencies, with both the /l/ and /k/ contingencies indicating that "ea" is likely to be the correct spelling.

TABLE 2

SEQUENTIAL CONTINGENCIES IN THE SPELLING OF /iy/

| Spelling  | Contingency         |     |     |     |      |     |     |             |     |     |      |
|-----------|---------------------|-----|-----|-----|------|-----|-----|-------------|-----|-----|------|
|           | Phoneme after /iy/* |     |     |     |      |     |     | Before /iy/ |     |     | None |
|           | /d/                 | /m/ | /k/ | /p/ | /s/  | /c/ | /h/ | /l/         | /b/ | /w/ |      |
| ea        | .20                 | .75 | .67 | .31 | 1.00 | .86 | .12 | .71         | .60 | .22 | .46  |
| ee        | .80                 | .17 | .33 | .69 | 0    | .14 | .44 | .29         | .30 | .67 | .42  |
| All Other | .05                 | .08 | 0   | 0   | 0    | 0   | .44 | 0           | .10 | .11 | .12  |

\* Based upon the assumption that occurrences of these phonemes are independent of the phonemes preceding /iy/. This is not strictly so; about 6 per cent of the relationship is attributable to other antecedents.



The improvement in spelling resulting from knowledge of contingencies as in TABLE 2 can be estimated by applying them to all of Thorndike's first 6000 monosyllables pronounced with /iy/. There are 171 such words, and the table is applicable to 58 per cent of them. The correct spelling of /iy/ is predicted for 77 per cent of these words, considerably better than the 46 per cent obtainable from single sound-letter correspondences. This sort of analysis should be extended not only to other sets of sound-letter relationships, but also to more representative samples of vocabulary to determine the extent of such useful contingencies in English.

Insofar as the above sorts of contingencies are an important feature of the within-word structure of English, and students can be taught to respond to them, the analyses of Horn and Moore are only a start on the type of data needed to design effective spelling programs. The job of analysis is large, however, even with high speed data processing equipment and not to be undertaken lightly. Such an undertaking was beyond the scope of this research.

Transfer of training and "generalization" are classical problems of spelling performance and bear a close relationship to the structure of English phonics. There are two distinct but related problems of transfer to be considered: spelling performance under writing conditions, and the influence of learning to spell specific words upon the spelling of other words. The latter of these cases has usually been considered under the label "generalization" in studies of spelling (Gates, 1935), but usage of the term is ambiguous, sometimes being descriptive and other times explanatory. The terms generalization and transfer will be used here in

a descriptive sense, to refer to performance under conditions which differ in some manner from the learning situation. It is believed that adequate explanation of transfer in spelling lies in a close examination of the nature of the subject-matter and the training procedures used.

Plentiful evidence is available to indicate that most spelling errors can be classified as "phonetic," that is, incorrect spellings which are pronounced the same as the correct spelling (Mendenhall, 1937, p. 10; Wolff, 1952). Mendenhall reported that about 75 percent of the errors he counted could be classified in this manner. The great bulk of these errors are attributable to the complex phonics of English. On the other hand, it is more difficult to obtain comparable information upon the degree to which words are correctly spelled through the regularity that exists in English phonics. Perhaps, Gates' research on Generalization and transfer in spelling (1935), gives as clear a picture as is available of the fact that students can be taught to transfer general rules for spelling to new groups of words, previously untaught. Since Gates was working with relatively regular aspects of English spelling, however, his research does not give indication of the potential usefulness of spelling rules and other devices to promote transfer in the more complex areas of spelling. Adequate information on the possibilities for positive transfer in English spelling awaits more complete analyses of English phonics.

There are several interesting hypotheses on the origins of negative transfer in spelling which deserve examination, however. If the phonics of English is inadequate to guide spelling, students will use the phonics they know, inadequately. If students are not adequately taught the phonics that



is known to be useful, they will also attempt to get along on what they know. Although no data has been gathered to support the notion, it is likely that the phonics training received by most students is not systematic, but incidental to other "language arts" activities, and so inadequate to the complexities of spelling. It is also likely that a great deal of the poor spelling which appears in the middle elementary grades and thereafter is negative transfer, generated by the selective nature of previous phonics training. Perhaps the greatest amount of phonics training received by children comes in connection with early reading instruction. Since this instruction is based upon "controlled vocabulary," any phonics based upon it is bound to provide an inadequate representation of the complexities of English spelling. Although a simplified phonics may be important in learning to read with some programs of instruction, it will lead to poor spelling unless supplemented by tasks which present a true picture of the necessary, complex phonics.

It is reasonable to expect that a spelling curriculum, represented by the corpus of spelling words to be taught, would be designed to give a valid picture of phonics, and to make up for the bias of early reading instruction. In practice, this does not appear to be the case. Word lists for spelling curricula are generally made up to meet criteria such as those of Horn (1950). Among the most important criteria are: frequency of use; wide geographic distribution; spread among various types of writing; quality of the words, in the sense of preferred usage; cruciality, in the sense of being important socially, economically, or otherwise; permanence in the vocabulary over time; and difficulty of spelling. The "Basic Life Spelling Vocabulary" of Fitzgerald (1951), represents one of the more modern and better of this sort of

word list. By means of comparisons between his list of 2650 words and more extensive counts of the words used in child and adult writing, Fitzgerald is able to state that "A typical child who mastered this list would be able to spell approximately 95 per cent of the running words he would need in his writing throughout his life (Fitzgerald, 1951, p. 54)." Neither this statement of utility nor the above criteria for word selection provide evidence that such a list will represent the spelling structure of English adequately enough to provide for some measure of positive transfer to the 5 per cent of running words not covered. It is also possible for such a list to contain too few instances of certain sound-letter relationships for adequate teaching of the more useful phonics in the basic vocabulary, itself.

The related problems of bias in early phonics training and in basic spelling vocabularies could be attacked by two procedures. One procedure would determine how adequate a sample of English phonics is provided by basic spelling vocabularies. The other procedure would devise a special teaching program designed to reduce negative transfer by extinguishing student dependence upon useless phonics as a basis for spelling certain words.

Since there is great overlap between basic spelling and reading vocabularies, especially for the elementary grades, a test for bias in spelling vocabularies would provide evidence not only about the adequacy of spelling vocabularies, but also about the notion that early reading instruction results in poor phonics training for spelling purposes. A definitive test of how well spelling lists represent the problems of English phonics is not possible because adequate analyses of the sound-letter relationships of English



are not available. However, a useful test might be made on the basis of letters alone. If sequences of letters which appeared in large samples of running text or in extensive word lists were either missing or greatly attenuated in spelling lists, there would be good reason to suppose that the phonics, of which the letters were a part, was also missing. Since such an analysis would be arduous, yet give no information about what phonics was missing and so contribute nothing to the construction of a better spelling list, it has not been undertaken here. Ideally, information on the adequacy of spelling lists should come from the comparative analysis of sound-letter relationships in the lists versus extensive groups of words such as those of Rinsland (1945) and Thorndike.

The experimental version of the 4th grade program was designed to teach students not to depend upon auditory cues for the spelling of certain families of words, and thereby prevent negative transfer effects for cases where guessing would prove fruitless. In a sense, this program was designed to test the feasibility of offsetting the bad effects of early phonics training by means of specific teaching. Forty-one special items were prepared to carry out this teaching. They were of two basic types: those requiring students to seek common letters in a group of words, and those asking Ss to judge which words looked most like the word to be spelled. (Examples of the two item types can be seen in APPENDIX A, on Sheet 27, taken from the experimental version of the 4th grade program.) Items were deliberately designed to require letter and word matching tasks which could be carried out only on the basis of the letters, often in the face of conflicting sound cues. For ex-

ample on the problem of transfer to writing. This is for reasons which will become apparent below.

ample, in item 24 the correct choice provides negative auditory cues while the incorrect choice provides positive auditory cues.

Item 24. Does wooden look more like

would  
could  
should

or

moon  
tooth  
pool

Words with various spellings of the sound /u/, as in "book," were chosen because data from Moore (1951) indicated that there was little basis for the use of auditory cues in their spelling. The various spellings of /u/ and percentages of occurrence are: "oo"-31, "u"-29, "ou"-29, and "o"-11. It was hypothesized that this sort of item would lead students to ignore auditory cues as a basis for spelling other words falling into the same sound and letter categories as those used in the treatment. The hypothesis was tested by means of a transfer test that presented untaught words taken from the same categories.

According to the often heard complaints from those who teach spelling and edit other peoples' writing, attaining transfer from formal instruction to actual writing is a severe problem. As the story goes, students who achieve 100 per cent on tests week after week produce less than perfect spelling on written work. It is almost certain that teaching machine programmers will have to face the somewhat accusatory questions of educators who fear that a new technique will diminish the seemingly meagre transfer obtainable with conventional teaching methods. Although the question of transfer is no more applicable to teaching machine programs than to other teaching techniques, it is important as a part of the ultimately necessary validation of any teaching process. It is interesting to note, however, that little research has been done on the problem of transfer to writing. This is for reasons which will become apparent below.



First of all, is there evidence of inadequate spelling on written work? Many counts have been made of the words misspelled in childrens' written work. Fitzgerald (1951, p. 143) has combined the results from a number of these studies into figures which show that the percentage of error in writing all words varies from about 11.9 in the second grade to 3.4 in the sixth. In a particularly well conceived study, Swenson and Caldwell (1948) present the percentage error and standard deviation of the percentage error in a count of 116,416 words written in letters by children in grades 4 through 12. These data are presented in TABLE 3. Both the percentage error and the standard deviation show reasonably orderly decrease from the 4th to 12th grade. Whether or not one calls a particular level of error excessive is an arbitrary matter. Since any misspelling is socially unacceptable, perhaps it is fair to judge that the figures below represent excessive misspelling, especially for students above the 6th grade level where formal spelling instruction often ceases.

TABLE 3

PERCENTAGE OF WORDS MISPELLED IN CHILDRENS' LETTERS, BY GRADE IN SCHOOL\*

| Mis-<br>spellings  | Grade in School |     |     |     |     |     |      |      |      |
|--------------------|-----------------|-----|-----|-----|-----|-----|------|------|------|
|                    | 4th             | 5th | 6th | 7th | 8th | 9th | 10th | 11th | 12th |
| Percentage Error   | 9.8             | 9.0 | 5.8 | 4.2 | 2.2 | 2.2 | 1.3  | .8   | .7   |
| Standard Deviation | 7.4             | 9.5 | 4.7 | 3.2 | 2.5 | 3.6 | 1.7  | 1.8  | .8   |

\*From Swenson and Caldwell (1948).

Assuming that the above figures represent an excessive level of misspelling, are the errors a matter of transfer from the learning to the performance situation? Such a question can be definitively answered only by determining whether students who meet spelling achievement criteria, subsequently, during writing, misspell words it is assumed they have learned. In the absence of specific information on this point, data on the so-called "spelling demons" provide a second-order estimate of the transfer problem. In his summary of spelling error data, Fitzgerald (1951) reports that 43.99 per cent of the errors made on the 6th grade level were on the 222 most frequently misspelled words. Since 219 of these words are in the most common 2000, and thus a part of the elementary spelling curriculum which should be learned by the 7th grade, misspellings of these "demons" may represent transfer problems of some sort. Looking back to the Swenson and Caldwell data for the 7th grade, where 4.2 per cent error was made, 43.99 per cent of this error, or 1.85 per cent may represent transfer errors. Whether or not this figure is sufficiently large to cause concern is, again, an arbitrary matter.

Making a brave assumption that the above figures indicate a real transfer problem, why should it exist, and what is the potential influence of teaching machine instruction in the nature and amount of transfer obtainable? It is believed that explanations of poor transfer to writing can be found in the nature of spelling behavior and in traditional teaching techniques and achievement criteria. From an examination of these factors it will be possible to hypothesize on the influence of teaching machine instruction.

The enormous complexity and irregularity of English phonics and the problems it generates for spelling have been discussed previously. In the



face of inadequate sound cues, a student probably responds on the basis of any other stimuli which are correlated with spelling responses. Positional and semantic contingencies, sequential contingencies of letters and sounds, and other subtle stimuli—including those provided by direct sound-letter relationships—all combine to serve as complex discriminative stimuli in the presence of which spelling responses will be emitted. When a spelling response is relatively uncertain, as it often is, the particular letters written are probably under control of a delicate balance of the many factors which make up the discriminative stimulus. That is, there is probably a very delicate balance of the many spelling cues present, making it possible for relatively subtle characteristics of the environment to exert significant influence upon the nature of the response made. Under these circumstances, it is not surprising that the composing of written material should produce negative transfer. For example, the following sentence fragments illustrate mistakes that might reasonably be made. In ". . . he went for forty days . . ." the misspelling of "forty" is quite reasonable, since there is a close semantic relationship between "four" and "forty" and none between "for" and "forty." Juxtaposition of the two words probably heightens the effect, especially if there is emphasis upon "forty" as being a long time. In ". . . mother an father . . .," the misspelling of "and" is a reasonable outcome of the poor pronunciation often given to familiar expressions; and in ". . . did receive the letter, beleive me . . ." the misspelling of "believe" could be generated by letter sequences and their pronunciation in the preceding words. Since innumerable such effects can be generated in the act of writing, it is impossible to devise any specific, preventive treatment.

The elimination of errors generated by writing depends upon spelling programs which will teach students to make the best possible use of all the spelling cues which are available, phonics, semantic, etc., in order to reduce sensitivity of spelling responses to the accidental effects of the stimuli produced by writing. However, any instructional procedure that requires spelling responses in a writing context should by chance expose students to some of the factors which generate errors and thus increase the possibility of positive transfer to future writing. For this reason it was hypothesized that the teaching machine program, where spelling responses were inserted into meaningful sentences, would produce transfer superior to the regular classroom procedures used in the control groups.

Other determinants of transfer to writing are to be found in traditional teaching-learning techniques and modes of achievement evaluation used for spelling. Essentially, spelling is learned by rote procedures in most all traditional programs. Typical instructions for spelling study tell a student to look at the word, attend to various features of its spelling, look away and spell it orally to himself, look at it again to see if he has spelled it correctly, and repeat the process a number of times while writing the word. In practice, this produces a set of massed trials on each word to be learned, with all but the first few spellings being correct as a result of immediate memory. In most classrooms the great bulk of this study takes place under threat of the weekly spelling test, perhaps in the hour or so just preceding the test. Students who follow the procedures described above can generally expect to do well on the test. As a matter of fact, the more perceptive students probably discover that the closer to test time their study time



is, the better they will do. It is clear that such study procedures do not allow the spelling of words to come under control of appropriate stimuli, but short-circuit the process by allowing either direct copying or immediate memory to provide the basis for responding. Spelling acquired under such circumstances can be expected to behave as does other rote learned material, with large memory decrements in a short period of time. Meanwhile, by using a test close upon the heels of this rote study, the false impression is gained that students have met an adequate spelling criterion. Insofar as such traditional classroom procedures do not result in correct spelling on written work it is less a matter of ineffective transfer than of inadequate initial learning. It was expected that the teaching machine instruction would result in more effective spelling on written work because it prevented the massed, rote memorization of spelling words by taking up different words on alternate items of the program.

Although the preceding factors may be responsible for some of the bad spelling in written work, the major influences probably lie with teachers and other agents of social control rather than with the spelling curriculum and procedures for teaching particular words. In another area of schooling, the now classic studies of Bagley (1905) and Ruediger (1908) well illustrated the effectiveness of teacher influence upon neatness in arithmetic and the wider effects of valuing neatness in general. In like manner, teachers probably can and do influence the quality of spelling in written work by setting standards of technical excellence and providing consequences which reinforce good spelling; the writing of social letters and the publication of written works are commonly used for this purpose. To the extent that

teachers do not carry out such activities one should not be surprised to find little transfer. The data of Swanson and Caldwell further suggest the teacher's importance in this respect. As shown in TABLE 3, the percentage of spelling error drops to .23 of its 4th grade value by the 8th grade and to .32 of its 8th grade value by the 12th grade. These two drops in percentage error are very close to the same value, and indicate that just about as much spelling progress was made in the years of high school as in the elementary school. Since the formal spelling instruction given in high school generally is not systematic and occupies little time, it is interesting to speculate on origins of the indicated gains. Although there is no general leveling-off in the acquisition of new vocabulary in the high school years (Smith, 1941) the gains may represent the fact that most writing is done with a relatively restricted spelling vocabulary that is taught in the elementary grades and learned to spelling perfection in the upper grades. The major way that the spelling of this vocabulary would be perfected is through its use in writing, and the major influences towards better spelling would have to be mediated by the teacher.

Of all the factors influencing transfer of spelling to writing the teacher effect is probably the most difficult to account for and control in experiments. In order to obtain adequate information on the validity of a spelling program through the measurement of transfer to writing, teacher effects would have to be controlled by careful regulation of the standards used for written work and by the use of large numbers of classrooms. The use of standardized writing tasks, designed to give maximum opportunity for the use of words taught, would also contribute to more stable measurements. Since such controls were



not feasible in this research it was expected that data obtained from samples of classroom writing would not give an adequate indication of transfer. The mere fact that it was impossible to examine more than a small amount of the student writing made it unlikely that the spelling words taught would appear frequently enough for a meaningful analysis. Nevertheless, counts were made of the spelling errors in comparable sets of written material from both of the 6th grade groups to gain an indication of any great differences which might appear.

#### Method

Subjects.--Students in the 2nd, 4th, and 6th grades in the cooperating elementary school served in both experimental and control groups. In the 2nd and 6th grades the subjects were split by class, with entire classrooms serving as experimental and control groups. In the 4th grade all Ss in both classrooms received teaching machine instruction, with different children using alternate versions of the program. Because of unavoidable shifts in school population the 2nd grade control class was of unusual constitution, having both 2nd and 3rd grade students. This meant that the 2nd graders in the class were accidentally exposed to 3rd grade curriculum throughout the year, giving them opportunity to learn subject-matter the 2nd grade experimental group never saw. The number of 2nd grade control Ss was reduced also.

Further descriptive information on the Ss is given in TABLE 4. It will be seen that the experimental and control groups are closely matched on the variables reported in TABLE 4. None of the differences between experimental

TABLE 4  
CHARACTERISTICS OF THE 2ND, 4TH, AND 6TH GRADE  
EXPERIMENTAL AND CONTROL GROUPS

| Characteristic of Group                                        | Group     |                  |           |     |           |     |
|----------------------------------------------------------------|-----------|------------------|-----------|-----|-----------|-----|
|                                                                | 2nd Grade |                  | 4th Grade |     | 6th Grade |     |
|                                                                | E         | C                | E         | C   | E         | C   |
| Number                                                         | 29        | 16               | 30        | 33  | 31        | 26  |
| Mean IQ Score                                                  | 114       | 113              | 112       | 111 | 115       | 116 |
| Mean Grade Equivalent Score<br>(less spelling) at End of Study | 3.5       | 3.8 <sup>a</sup> | 5.9       | 6.0 | 7.8       | 8.1 |
| Mean Gain in Score<br>(less spelling) During Study             | b         | b                | .6        | .6  | 1.0       | 1.1 |

<sup>a</sup>Significant at .01 level (Mann-Whitney test, one-tailed).

<sup>b</sup>Pretest data not available (see text).

and control groups approach significance except that of the 2nd grade mean achievement score at the end of the experiment. The difference here favors the control group, and since the scores show combined achievement for all areas except spelling, they probably reflect the effects of the mixed 2nd and 3rd grade character of the class. It was impossible to compare the 2nd grade groups on average achievement gains during the year because no achievement data was available in the school for students prior to the 2nd grade. However, it is likely that the control group would show the greater gain on such a comparison. Further data on the matching of control and experimental groups will be presented when the results of particular treatments are discussed.



Measures.--The IQ scores reported in TABLE 4 were those on file in the records of the cooperating school. Except for individuals who had been given individual tests, the scores were derived from Otis Quick-Scoring Mental Ability Tests. Stanford Achievement Test Batteries administered by the school in the normal course of student testing provided the data in TABLE 4 on achievement status of the students used in the experimentation. The mean grade equivalent scores (less spelling) were obtained from the tests administered at the end of the school year during which the experimentation took place, and mean gains in grade equivalent score were derived from the differences between these tests and those of the previous spring. The spelling subtests of the Stanford Achievement Batteries provided measures of the experimental treatment parallel to other tests to be described below.

A standard measure of spelling achievement independent of the normal testing program in the school was provided by the spelling subtests of the primary and elementary batteries of the California Achievement Tests. In addition to these standardized tests, other measures were made of student gains in the specific context of the spelling programs. Weekly spelling tests that covered all of the words taught were administered by the cooperating teachers to students in the 2nd and 6th grade experimental and control groups. In the 4th grade experiment a 43 item test which systematically sampled the 70 words taught was administered as both a pretest and posttest. This test and the weekly tests were dictated tests of the conventional sort.

Three special instruments were designed for the studies. The 6th grade experimental group was administered a questionnaire designed to tap their reactions to teaching machine instruction. A detailed description of this

instrument is given in the section of this report on experimental results. A word matching test, patterned after items described by Gates (1928, p. 47), was constructed to match students for experimental treatment. The 60 items in the test were of one basic type, with equal numbers of items using English words, 3rd order, and zero order approximations to English words. Item types appear below. The task is to select from among the various alternatives the one word that exactly matches the given sample. It was felt that this test

|          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|
| fourteen | fuorteen | fonrteen | fcurfeen | fcnrtean | foorteen |
|          | fuurteen | feurteen | foarten  | fourteen | fnurteen |
| plaknort | plalnort | plahnort | plkanort | placnort | plagnort |
|          | plankort | plafnort | plaxnort | plakkort | plaknort |
| uodjujy  | udojujy  | uodjujy  | ucjdujy  | uedjujy  | udojujy  |
|          | udjujy   | uedjujy  | udejujy  | ucbjujy  | ubpjujy  |

would give information on how adequately individuals made the matching responses necessary in confirming their answers to the teaching machine items. A transfer test of the dictated variety was designed to test pervasive effects of the experimental 4th grade program. Words on this test were taken from the same sound and letter categories as words explicitly taught in the program, so that there was some basis for expecting transfer to take place.

Handwriting of the subjects was judged from samples of their answers to testing items rather than from free composition. This procedure was expected to lead to relatively standardized samples that would be representative of the style and care of writing used in studying both with and without the teaching machines. The samples were scaled by matching with Scale A of Thorndike's Handwriting Scale (Thorndike, 1917), and fell into Qualities 5



through 13. Two raters judged each sample, and had little difficulty in reaching agreement. Correlation between ratings of the two judges was .96 ( $N = 62$ , with Spearman rank correlation corrected for ties).

Measures of error rate, cheating, and misscoring were derived from analysis of student responses to the teaching materials. Students had only one opportunity to respond to each teaching item, and each sheet of five items dropped into the far end of the teaching machine after use. These used sheets were picked up weekly by the experimenters for analysis. Errors were recorded by student and by teaching item, with individual student records showing errors per sheet of five items. Since Ss marked each of their responses to a teaching item as correct or incorrect by means of a "+" or a "0," misscorings were counted by noting disagreement between scoring and response to the items. The most prevalent form of cheating involved pulling an item back into position in the response window after having seen that it was wrong. This could be done only by means of considerable force, which invariably left marks from the student's fingers and from the paper drive mechanism. Instances of this evidence in the response protocols provided data on cheating. The only other form of cheating possible was to answer all of the items before placing sheets in the teaching machines. No cheating of this sort was observed, perhaps because all use of the machines took place in supervised classrooms.

Procedure.--In the 4th grade experiment teaching machines were available for one-third of each classroom group at a time, with the machines being shared by the two classrooms. Use of the machines was scheduled to fit the

other requirements of normal classroom work. Each student had a packet of the appropriate version of the program from which the necessary material could be taken as needed. No other normal spelling was taught for the one-month duration of the experiment, and the same procedures were used throughout.

Procedures in the 2nd and 6th grade experiments were more complex because of non-teaching machine methods in the control groups and because procedures were altered as the experiments progressed. The control groups followed the methods outlined in the normal text materials, which turned through a 5 day cycle. Both the 2nd and 6th grade control groups used these procedures.

On the first day a story was read and questions answered on its content. This was followed by individual study of the new spelling words for the week.

The second day students did specific exercises prescribed in the book.

On the next day a test was dictated, followed by self-scoring and study of the words missed.

The fourth day involved further book exercises and independent study of words that individuals needed to emphasize.

The last day of the cycle included another test, which the students corrected. The missed words were entered into a personal notebook.

Sufficient teaching machines were available for one-half of each experimental group, and this number seemed to present no scheduling difficulties. Teaching materials in current use were stored in shelves readily accessible to students as needed. Initially, teaching machine instruction took place in the usual spelling period, but as the year progressed it was found convenient to use the machines when there was a break in the usual schedule. This increased the chance of students having to stop work part way through a sec-



tion of the program, which was felt to be undesirable, but students were instructed to go back several items when they picked-up the work again in order to reorient themselves to the task.

In the 6th grade experimental group, the first 9 lessons of the year and lessons 17, 25, and 34 were taught by the teacher in the usual fashion. Lessons 10 through 16, 18 through 24, and 26 through 33 were taught by teaching machine. At the beginning of lesson 19 the teacher told the class that they should attempt to reduce the number of errors made in doing the teaching machine lessons, and that programmed sheets on which errors were made would have to be done over again the next week. During the weeks of lessons 20 and 21, a class list was posted that told what sheets from the preceding week would have to be done again, and Ss were required to check their names off when the task had been done. Beginning with lesson 22 and continuing for the balance of the year, a sheet with the following directions was posted for the lesson of each week: "Check off each spelling sheet only after you have answered every question on it correctly. If you miss any questions on the sheet, do the same sheet over again. When you have done all of the sheets correctly, then check the X column" (signifying completion of the task).

In the 2nd grade experimental group, four distinct procedures were followed, dictated in part by the age and lack of sophistication of the youngsters. For lessons 1 through 6 the teacher taught spelling as usual, and in addition worked with groups of children in showing them how to use the teaching machines. During lessons 7 through 11 children worked independently with the teaching machines if they were able to do so, and individuals who had difficulties with reading the programs and completing them properly were helped by the

teacher. Because most of the difficulties were encountered in individuals who were not able to read at the required 2nd grade level, a set of magnetic disk recordings was prepared for lessons 12 through 17 to exactly parallel the text in the teaching machine program. Students with reading problems used the recordings by playing them item by item as they worked at the spelling program. Lessons 18 through 22 were taught by the teacher without any teaching machine instruction. For lessons 23 through 34 a set of self-administering tests were prepared on magnetic disks, one for each lesson. After having finished a section of the program, a student would give himself the test, have it corrected by the teacher, and either go on to the next lesson or repeat the same lesson depending upon whether the test was perfect or not. Students were permitted to proceed through the spelling program as rapidly as they desired, constrained only by the time available and limited physical facilities.

### Results and Discussion

Results of the teaching machine instruction are reported here along with further information on English spelling. Findings which bear upon each of the previously discussed problems are given and discussed in the same order as under The Problems Studied, pages 34-64. This is followed by a more general discussion.

Teaching efficiency.--Estimates of the effectiveness of the teaching machine treatments are available for both 2nd and 6th grades. TABLE 5 presents median pretest and posttest scores for the California Achievement



Tests, the Stanford Achievement tests, and combined scores on the 6th grade level. In all groups there were gains in grade equivalent scores during

TABLE 5

COMPARISON OF MEDIAN GRADE EQUIVALENT SCORES OF 2nd AND 6th GRADE  
EXPERIMENTAL AND CONTROL GROUPS BEFORE AND AFTER TREATMENT

| Median Observed Scores |   |                 |           |       |               |           |       |                |           |       |    |
|------------------------|---|-----------------|-----------|-------|---------------|-----------|-------|----------------|-----------|-------|----|
| Group                  |   | California Test |           |       | Stanford Test |           |       | Combined Tests |           |       | N  |
|                        |   | Pre-test        | Post-test | Diff. | Pre-test      | Post-test | Diff. | Pre-test       | Post-test | Diff. |    |
| 2nd Grade              | E | 3.3             | 4.0       | .7    | a             | 3.6       | --    | --             | --        | --    | 29 |
|                        | C | 3.3             | 4.4       | 1.1   | a             | 4.2       | --    | --             | --        | --    | 16 |
| 6th Grade              | E | 6.5             | 7.0       | .5    | 6.0           | 7.8       | 1.8   | 6.2            | 7.3       | 1.1   | 31 |
|                        | C | 6.8             | 7.5       | .7    | 6.4           | 7.9       | 1.5   | 6.7            | 7.8       | 1.1   | 27 |

Medians of Differences Between Observed Pretest and Posttest Scores

| Measure         | Group     |     |       |                |           |     |       |                |
|-----------------|-----------|-----|-------|----------------|-----------|-----|-------|----------------|
|                 | 2nd Grade |     |       |                | 6th Grade |     |       |                |
|                 | E         | C   | Diff. | p <sup>b</sup> | E         | C   | Diff. | p <sup>b</sup> |
| California Test | .9        | 1.1 | -.2   | <.005          | .4        | .7  | -.3   | .08            |
| Stanford Test   | a         | a   | --    | --             | 1.4       | 1.1 | .3    | <.02           |
| Combined Tests  | --        | --  | --    | --             | .9        | .7  | .2    | .12            |

<sup>a</sup>Stanford pretest data unavailable for 2nd grade.

<sup>b</sup>Mann-Whitney U Test, single-tailed.

the year, with gains being greater on the Stanford than on the California achievement tests. For both the 2nd and 6th grades the control groups show

greater gains than the experimental groups on the California test, while the 6th grade experimental group shows greater gains than the control group on the Stanford test. Combining the California and Stanford scores for the 6th grade results in approximately equivalent gains for both groups.

Effects of the teaching treatments were analyzed for statistical significance by comparing the distributions of gains for corresponding experimental and control groups. That is, pretest to posttest differences in observed scores were calculated for each group, and differences between these distributions of appropriate experimental and control pairs were tested for significance by the Mann-Whitney U Test (Siegel, 1956). Medians of these distributions of differences between pretest and posttest scores are given in the lower half of TABLE 5 along with their differences and significance figures. It should be noted, for the acute observer, that these medians of the difference distributions do not correspond exactly to the differences between pretest and posttest medians. This is because the median, a positional measure, is sensitive to changes in rank order of individual observed scores from pretest to posttest.

For the 6th grade data the Stanford test shows a significant difference in favor of the experimental group, the California test shows a difference just short of usual standards of significance in favor of the control group, and the combined scores a slight but not significant difference in favor of the experimental group. No striking claims can be made for the teaching machine treatment on the basis of this evidence. It is interesting to speculate why the California Test, administered especially for this study, did not show as large gains as the Stanford. This test uses a multiple choice format with



subjects required to indicate which of four words is misspelled, if any, while the Stanford test requires the writing of words presented in sentence contexts. The latter test more nearly parallels the procedures used in the teaching machine treatment. Possible motivational factors will be examined in the section on year-long changes in performance.

The 2nd grade data indicate a highly significant difference in favor of the control group. Posttest results on the spelling section of the Stanford Test, reported in TABLE 5, which show a difference of .6 grade equivalent score in favor of the control group (significant at the .002 level; Mann-Whitney U test, single-tailed), also support the notion of a real difference, but adequate interpretation of these findings depends upon an examination of further information. Because the 2nd grade control subjects were members of a mixed 2nd - 3rd grade class, they were exposed to 3rd grade spelling materials, and according to the control group teacher they tried to spell the 3rd grade words along with their older classmates. It may be that the gains favoring the control group were based entirely upon this experience, which exposed students to advanced portions of the curriculum, and not upon the techniques involved. This hypothesis was tested by rescoring the California spelling tests on the basis of letters correctly spelled in the dictated words rather than on the basis of entire words. Each correct letter in its correct position in a word was given one point, and points were subtracted for both wrong letters and for correct letters in the wrong positions. It was thought that this procedure would give subjects credit for partly correct words which they had never been exposed to in the curriculum and, in a sense, partially counteract the effects of special exposure received by the control group.

Insofar as the basic behavioral units in spelling are sequences of letters rather than entire words, this is a valid scoring procedure. Results of the scoring and its effects upon the comparison of the 2nd grade groups are shown in TABLE 6. Findings with this technique of scoring reverse the previous

TABLE 6

COMPARISON OF MEDIAN SPELLING GAINS OF 2nd GRADE GROUPS,  
BASED UPON LETTER SCORING OF THE CALIFORNIA TEST

| Median Observed Scores                         |         |          |       |
|------------------------------------------------|---------|----------|-------|
| Group                                          | Pretest | Posttest | Diff. |
| E                                              | 78      | 144      | 66    |
| C                                              | 122     | 156      | 34    |
| Medians of Differences Between Observed Scores |         |          |       |
| Group                                          |         | Diff.    | P*    |
| E                                              | C       |          |       |
| 63                                             | 33      | 30       | .001  |

\*Mann-Whitney U test, single-tailed, N=29,16.

results and give advantage to the teaching machine treatment. The practical significance of this finding depends upon validity of the letter scoring technique, whether letter scoring correlates highly with more conventional word scoring which is the social criterion of good spelling.

Further indications of the relative effectiveness of the teaching machine treatment are available from an analysis of weekly test scores. In the 6th grade two sorts of comparisons are possible, between matched-lesson test scores of the experimental and control groups and between teacher and machine



lessons in the experimental group. Similar comparisons would not be meaningful in the 2nd grade because of the complex procedures followed in the experimental group.

Mean weekly test scores across subjects for each of the 22 machine taught lessons were calculated for the experimental group and for the corresponding lessons in the control group. Results show very little range, with all scores falling near the maximum of 17 words. For the experimental group the mean is 16.33 with a range of 15.73 to 16.86, while the control group received a mean of 16.64 with a range of 15.98 to 16.95 (the range for individual subjects is the same for both groups, 10.03 to 17). The difference between means was tested by the Wilcoxon matched-pairs test, matched on the 22 lessons. With  $N=22$  and a  $T$  of 34.0, the difference in favor of the control group was significant at the .005 level for a single-tailed test. This result appears to have been caused by factors other than the machine versus teacher treatment, however. The mean weekly test scores for the 11 non-machine (teacher) lessons in the experimental group produced a mean of 16.29, less than the machine taught lessons, indicating that some factors in the teaching or testing situations in the two classrooms other than the machine treatment were responsible for the difference. From direct observations made in the classrooms, the control teacher encouraged and gave more opportunity for last minute study before the weekly posttest, and this could have produced the difference. In the control group students corrected their own tests just after having taken them, so that the consequences of any poor spelling were immediately apparent. This probably had a positive effect upon test performance which was not present in the control group where the teacher graded

tests and handed them back a day or so later. Whether these are adequate explanations of the differences reported is not known, but it is apparent that a meaningful comparison cannot be made between the experimental and control groups on the basis of weekly test scores.

Comparisons between machine and teacher taught lessons in the experimental group had the benefit of each subject serving as his own control and of no inter-classroom differences. The mean weekly test score of all teacher lessons is 16.29 words correct, and the mean of all machine lessons is 16.35 words. Using the Wilcoxon matched-pairs test, matched on students with  $N=29$ , the  $T$  was 199 and the  $P=.34$ . Evidently it is impossible to make any claim for either method of teaching from this over-all comparison. This is not surprising in view of the over-all insensitive nature of the weekly spelling tests. The fact that the mean weekly spelling test performance is so high not only produces an insensitive spelling measure, but also casts doubt upon the validity of the measure as a criterion of adequate spelling performance. In the next section of this report where year-long changes in performance are discussed, the achievement during particular sections of the year will be examined and some interesting differences will be noted.

Although there is a trend in the 6th grade results towards superiority of the teaching machine group and in the 2nd grade there are highly significant results if the letter scoring measure is valid, no striking and consistent findings have been produced in favor of the teaching machine treatment. This picture is changed drastically by considering learning efficiency rather than straight learning outcomes. By direct observation, the normal time spent on spelling instruction in both the 2nd and 6th grades



was approximately 20 minutes per day, with periods of up to 30 minutes on occasion. Of this time, approximately 20 minutes was used for two spelling tests, allowing some 80+ minutes of spelling study-time a week. For the teaching machine programs, average study time per week was approximately 25 minutes, one-third of the normal time. The effect of this saving in the 6th grade experimental group was to decrease the amount of time given to

TABLE 7

NUMBER OF STUDENTS PER WEEK FINISHING LESSONS 23-34  
ON AN AD LIB. BASIS

| Week | Lesson |    |    |    |    |    |    |    |    |    |    |    |
|------|--------|----|----|----|----|----|----|----|----|----|----|----|
|      | 23     | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 |
| 1    | 31*    | 25 | 20 | 16 |    |    |    |    |    |    |    |    |
| 2    |        | 31 | 27 | 24 | 18 | 16 | 11 | 8  |    |    |    |    |
| 3    |        |    | 30 | 26 | 22 | 19 | 13 | 10 | 4  | 3  |    |    |
| 4    |        |    |    | 31 | 23 | 21 | 15 | 12 | 7  | 7  | 5  | 3  |
| 5    |        |    |    |    | 30 | 25 | 22 | 16 | 15 | 11 | 7  | 7  |
| 6    |        |    |    |    | 31 | 30 | 24 | 21 | 16 | 12 | 12 | 11 |
| 7    |        |    |    |    |    | 31 | 30 | 22 | 19 | 15 | 15 | 15 |
| 8    |        |    |    |    |    |    | 31 | 24 | 22 | 21 | 20 | 17 |
| 9    |        |    |    |    |    |    |    | 28 | 26 | 22 | 21 | 20 |
| 10   |        |    |    |    |    |    |    | 29 | 28 | 26 | 26 | 24 |

\*The total number of students in class was 31.

spelling. In the 2nd grade, the variety and shift of student groupings and individual arrangements made it impossible to obtain an adequate picture of time factors until the latter part of the year when students were allowed to proceed through lessons 23-34 ad lib. A record of individual progress through these lessons was kept, and is presented in TABLE 7.

At the end of the first week of ad lib. spelling instruction half of the class had finished 4 lessons instead of the customary one, and after four weeks the number of lessons finished was 7, with a range of 4 to 13. After this point there was no more room for the range to spread because the teaching materials had been exhausted. In addition, the self-testing and test scoring facilities were unable to keep up with student demand so that the rate and range of material use was further hindered. Under these conditions the median rate of progress through the spelling lessons was approximately two times normal, and under less constrained conditions would have been greater.

Both the 6th grade measure of study time per lesson and the 2nd grade measure of lessons per unit time indicate that most students using the teaching machine programs under ideal conditions should be able to complete three times as much spelling instruction as under normal classroom procedures with equivalent learning outcomes. This finding is based upon crude measures and programs which may not be typical of other self-teaching materials, but it suggests, along with the achievement gains recorded by the mixed 2nd-3rd grade control group, that one of the greatest benefits of teaching machine instruction certainly derives from allowing students individual, free access to more advanced portions of the curriculum as soon as they are ready for it.



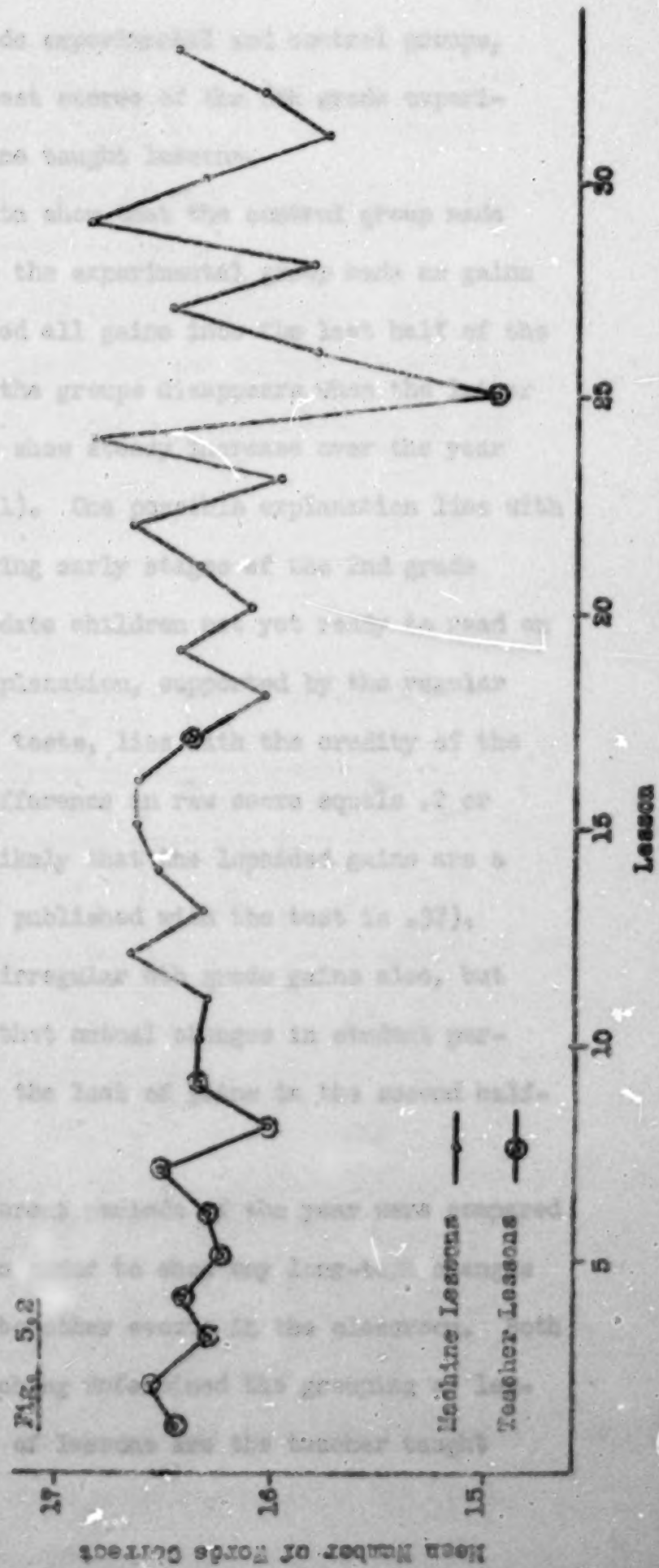
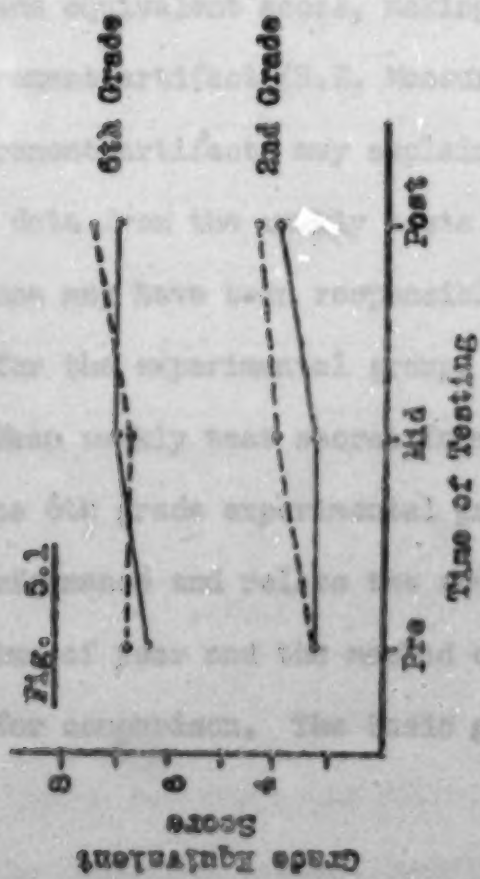
The hypothesis that the teaching machine treatment would result in more uniform learning outcomes than normal classroom procedures was not confirmed. Pretest and posttest variances for the experimental and control groups were computed and compared by means of the F test. For the 6th grade groups Stanford test data was used because it showed the greater amount of pretest to posttest change. For the 2nd grade groups California test data was used because Stanford pretests were unavailable. Pretest variances in the 2nd grade experimental and control groups were .37 and .40 respectively, indicating no initial differences between the groups. Posttest variances were .58 and .18, giving a variance ratio of 3.22 which is just significant at the .01 level with 27 and 15 degrees of freedom. Apparently the machine treatment raised variability of performance while the teacher treatment decreased it. Pretest variances in the 6th grade experimental and control groups were 1.9 and 3.2 respectively, not significant ( $P > .05$ ) with 30 and 21 degrees of freedom. Posttest variances were 1.7 and 3.7 giving a variance ratio of 2.18 which is significant beyond the .05 level with 26 and 30 degrees of freedom. Here the effects of machine treatment were opposite to the 2nd grade, producing changes in the direction hypothesized, but not large enough to be of great importance, if significant. The above results will be discussed in the section of this report that considers individual variables.

Year-long changes in performance.--The two major sources of information on year-long changes in performance are the California Spelling Achievement Test which was administered before, half-way through, and after the experiment, and the weekly posttests dictated in each classroom. The data to be examined are presented in FIGURE 5. FIGURE 5.1 shows results of the California Test

FIGURE 5: YEAR-LONG CHANGES IN PERFORMANCE OF STUDENTS TAUGHT BY TEACHING MACHINE

5.1 Pre- Mid- & Posttest Scores of 2nd & 6th Grade Experimental Groups on the California Spelling Test.

5.2 Mean Weekly Posttest Scores of the 6th Grade Experimental Group.





administrations to the 2nd and 6th grade experimental and control groups, and FIGURE 5.2 shows the mean weekly test scores of the 6th grade experimental group for all teacher and machine taught lessons.

The 2nd grade achievement test data show that the control group made steady gains throughout the year while the experimental group made no gains from pretest to midtest but concentrated all gains into the last half of the experiment. This distinction between the groups disappears when the letter scoring technique is used; both groups show steady increase over the year (these data are not plotted in Fig. 5.1). One possible explanation lies with the many changes in procedure made during early stages of the 2nd grade experimentation in attempts to accommodate children not yet ready to read on the 2nd grade level. A more likely explanation, supported by the regular gains shown with letter scoring of the tests, lies with the crudity of the measuring instrument where one word difference in raw score equals .2 or .3 grade equivalent score, making it likely that the lopsided gains are a measurement artifact (S.E. Measurement published with the test is .37). Measurement artifacts may explain the irregular 6th grade gains also, but other data from the weekly tests show that actual changes in student performance may have been responsible for the lack of gains in the second half-year for the experimental group.

Mean weekly test scores from different periods of the year were compared for the 6th grade experimental group in order to show any long-term changes in performance and relate the changes to other events in the classroom. Both the time of year and the method of teaching determined the grouping of lessons for comparison. The basic groups of lessons are the teacher taught

lessons (1-9), the early machine taught lessons (10-16), and the late machine taught lessons (22-33). Machine lessons 18-21 were excluded from analysis because they covered a period during which classroom procedures were in flux. Lessons 17 and 25 were excluded from analysis because they were isolated teacher lessons and lesson 25 is obviously off-scale. Comparisons between the first and second half of the year are based upon the combination of lessons 1-9 and 10-16 versus lessons 22-33. TABLE 8 presents results of the comparisons.

TABLE 8  
COMPARISON OF MEAN WEEKLY TEST SCORES OF 6th GRADE EXPERIMENTAL GROUP FOR DIFFERENT PARTS OF THE YEAR

| Lesson | Mean Score | Diff. | N  | T     | P*  |
|--------|------------|-------|----|-------|-----|
| 1-16   | 16.43      | .16   | 28 | 130   | .05 |
| 22-33  | 16.27      |       |    |       |     |
| 1-9    | 16.38      | .12   | 28 | 127   | .04 |
| 10-16  | 16.52      |       |    |       |     |
| 10-16  | 16.52      | .25   | 30 | 114.5 | .01 |
| 22-33  | 16.27      |       |    |       |     |

\*Wilcoxon matched-pairs test, single-tailed, matched on Ss.

Though the over-all comparison of lessons 1-16 in the first half-year with lessons 22-33 in the second half-year shows a difference of only .16, or a hardly practical 5 words over the 34 lessons of the year, the difference is significant and the other differences shown consistently favor the first half-year. Both the teacher lessons (1-9) and the machine lessons (10-16)



in the first half-year show significantly greater learning outcomes than lessons 22-33. Within the first half-year, lessons 10-16 show significantly greater outcomes than lessons 1-9, but this probably reflects the machine versus teacher comparison rather than a general rising performance within the first half-year. This comparison of adjacent machine and teacher lessons is probably the most valid of all the comparisons made because of roughly comparable classroom conditions throughout the entire 16 weeks.

It is more difficult to decide whether the above fall-off in achievement is a result of teaching machine instruction or a general effect of the year in school. In the absence of any section of teacher-taught lessons in the second half-year a comparison of the first to second half-year fall-off in teacher versus machine lessons is not possible. Estimates can be arrived at, however, by observing trends in the data. By inspection of Fig. 5.2 there appears to be a negative slope to the teacher lessons 1-9, and hardly any slope to machine lessons 10-16 and 22-33. Computing Spearman rank order correlations between the order of lessons and ranked scores of the lessons gives  $-.579$  for lessons 1-9,  $.375$  for lessons 10-16, and  $-.207$  for lessons 22-33. These results tend to support visual inspection of Fig. 5.2, with the correlation of  $-.579$  for lessons 1-9 being the only one approaching significance (for a  $P$  of  $.05$ ,  $\rho$  must equal  $.600$ ). Using Fisher's method (Vert, Neidt, and Ahmann, 1954) of testing for the significance of the differences between the correlation coefficients shows only lessons 1-9 and 10-16 to differ ( $P=.05$ ). This indicates significantly different trends in the teacher versus machine lessons in the first half-year, with the teacher lessons moving towards less achievement more rapidly than the machine lessons.

In view of the negative correlation for lessons 22-33 it is probably unwise to extrapolate this finding into the second half-year, but it is highly unlikely that the machine lessons would show any greater fall-off than the teacher lessons.

The above findings refute the notion that teaching machine instruction is a procedure that cashes-in on short-term motivational effects of a novel situation. Rather, it has been found to behave much like one other, more traditional instructional procedure over the year of experimentation. Further support for this conclusion comes from a failure to find any enhancement of achievement on lessons 10, 14, 18, 26, and 28 which were the first lessons taught by machine after teacher lessons or vacations. If there had been strong motivational effects from the novelty aspects of teaching machine instruction they should have been increased by the periods away from teaching machine instruction and been reflected in greater achievement upon renewed exposure to the technique.

Individual variables.--Results already presented have shown that the 6th grade experimental group gained significantly more than the control group when the results were measured on the spelling section of the Stanford Achievement Test. In addition, there was some indication of a decrease in variance in the experimental group along with an increase in the control group. These results will be examined in more detail.

TABLE 9 presents these same achievement results when the groups are split into upper and lower halves at the median IQ scores. There is no difference between the upper experimental and control groups, but a large and significant difference between the lower groups, with the experimental group



TABLE 9  
COMPARISON OF MEDIAN GAINS OF THE 6th GRADE GROUPS  
SPLIT INTO UPPER AND LOWER HALVES AT THE  
MEDIAN IQ SCORES OF EACH

| Group |   | Median Scores |                  |                | Diff. | N  | P*   |
|-------|---|---------------|------------------|----------------|-------|----|------|
|       |   | IQ            | Stanford Pretest | Stanford Gains |       |    |      |
| Upper | E | 122           | 6.2              | 1.20           | 0     | 28 | n.s. |
|       | C | 122           | 7.2              | 1.20           |       |    |      |
| Lower | E | 107           | 5.8              | 1.55           | 1.05  | 29 | .006 |
|       | C | 109           | 5.5              | .50            |       |    |      |

\*Mann-Whitney U test, single-tailed.

showing greater gains. Since the groups are similar in IQ score, this does not explain the result, nor does the slight difference in Stanford pretest scores, because the correlations ( $\rho$ ) between these scores and the Stanford gains are  $-.168$  and  $-.029$  respectively, for the experimental and control groups. A similar comparison between upper and lower IQ halves within the experimental group shows the high IQ students made a median gain of 1.20 while the low IQ students made a gain of 1.55 grade equivalent score. This difference is not significant. In the control group the high IQ gain is 1.20 while the low IQ gain is .50, a difference significant at the .03 level (Mann-Whitney U test,  $N=13,13$ ). Results similar to these but not so striking are obtained if the groups are split on the Stanford pretest scores.

These results very strongly support the hypothesis that teaching machine instruction produces increased gains for individuals in the lower half of the

IQ range, but there are related consequences of the teaching machine treatment that will become clear through an examination of the correlations between gains on the Stanford test and measures of individual differences. TABLE 10 shows the rank order correlations between Stanford gains and IQ and Stanford pretest scores for the total experimental and control groups and for the upper and lower halves of the groups when split on IQ scores. Concentrating first on the total group, the IQ correlations with gains are  $-.128$  for the

TABLE 10

RANK ORDER CORRELATIONS OF GAINS ON THE STANFORD SPELLING TEST VS. IQ AND STANFORD PRETEST SCORES FOR UPPER AND LOWER IQ HALVES AND TOTAL 6th GRADE EXPERIMENTAL AND CONTROL GROUPS

| Variable | Group             |                    |                    |         |       |                   |
|----------|-------------------|--------------------|--------------------|---------|-------|-------------------|
|          | Experimental      |                    |                    | Control |       |                   |
|          | Lower             | Upper              | Total              | Lower   | Upper | Total             |
| IQ       | .681 <sup>a</sup> | -.517 <sup>b</sup> | -.128              | .236    | -.014 | .343 <sup>b</sup> |
| Pretest  | -.168             | -.407              | -.538 <sup>a</sup> | -.029   | -.415 | -.061             |

<sup>a</sup>Significant at the .01 level ( $N_{\text{lower}} = 16$ ,  $N_{\text{total}} = 31$ ).

<sup>b</sup>Significant at the .05 level ( $N_{\text{upper}} = 15$ ,  $N_{\text{total}} = 26$ ).

experimental group and .343 for the control group, a significant correlation. This corresponds to the finding one would expect from the differential gains of the upper and lower IQ groups reported in TABLE 9, but correlations within the upper and lower halves of the IQ distribution reveal a curvilinear relationship showing that these total correlations are as much a matter of what happens to the upper as to the lower groups. IQ scores in the lower experimental group correlate .681 with gains, while in the upper group the correlation is  $-.517$ .



Both correlations are significant. In the control group lower and upper correlations are .236 and -.014, respectively. Neither correlation is significant.

The interpretation of these results is complicated by the correlations between gains and the spelling pretest, but can be teased-out with patience. Within both the experimental and control groups there is negative relationship between spelling pretest scores and gains, with the relationship being stronger for the upper IQ groups. This negative relationship probably reflects the fact that when a subject-matter (or a test) has a ceiling, the more a student already knows the less gain he will be able to show. For the low IQ students the ceiling was farther away, allowing more potential gain and lowering the negative relationship. Insofar as the correlations between IQ scores and gains reported in TABLE 10 are not independent of the pretest scores, the same ceiling effect was probably partly responsible for the IQ to gain relationships shown. On the assumption that the rho's in TABLE 10 approximate the Pearsonian r's, partial correlations between IQ scores and gains, independent of the pretest scores, were calculated. For the lower and upper sections of the experimental group they are .670 and -.461; for the control group they are .286 and .086. These figures are all close to the original first-order correlations, indicating that most of the relationship between IQ and gains must be explained by other means than the ceiling effect of limited subject-matter.

The high positive correlation between IQ and gains in the lower IQ experimental group shows that when there was a high ceiling on the subject-matter students taught by the teaching machines tended to increase in

individual differences in achievement. In the control group on the other hand, there was also no low ceiling, but the teaching techniques did not produce much in the way of learning outcomes and so did not allow the inherent individual differences to develop.

The essentially zero correlation between IQ and gains in the high IQ control group was probably generated by the relatively narrow range of achievement possible for this group. With a pretest grade equivalent score of 7.2 and 6th grade materials to study, there was hardly any chance for improvement and the improvement made was evidently well within the grasp of all members of the group. In the experimental group on the other hand, the pretest score was 6.2, indicating more room for gains, but the correlation between gains and IQ is significantly negative. Since this correlation is not a function of the pretest status of the students, it is apparent that individuals who had room for improvement and presumably the greater "ability" for improvement did not improve up to expectation. The students in this group studied the teaching machine program more "properly," on the average, than the students in the low IQ group, with the median frequencies of cheating, misscoring, and errors being 6, 9, and 40 as opposed to 8, 24, and 78 for the low group, so that inadequate study procedure by the group does not explain the negative relationship. A reasonable explanation may lie with the type of reinforcement provided by the program. In the absence of any intellectually stimulating content, the knowledge of results provided by the program was probably reinforcing only in the sense of informing a student that he had made the correct response. Although knowledge of success may be a reinforcing event for all students, it is probably more reinforcing for



the poorer students than for high IQ students who are generally successful even when studying poorly programmed ("difficult" and "challenging") lessons under conditions of intermittent knowledge of results. The program was probably accepted by these students as a necessary nuisance imposed by the teacher and completed with competent indifference.

Apparently teaching machine instruction does not destroy the usual relationship between IQ scores and achievement but can enhance it under some conditions when the choice of teaching material matches the level of the students and the programming procedures are appropriate. The notion that teaching machine instruction may increase the predictive power of certain other individual variables was tested on the 2nd, 4th, and 6th grade levels for the word matching test previously described. Results indicate rho's of about .250 between the word test and spelling gains for every group, including the lower IQ group in the 6th grade where the greatest gains were recorded and where the test might be expected to correlate most highly. It was also expected that the test would predict better on the 2nd grade level where students still learning to read are not so skillful at word matching tasks. This was not the case. None of the correlations approach significance. The explanation probably lies in the fact that on all of the grade levels involved the number of misscorings was very small. On the 2nd, 4th, and 6th grade levels 5.8, 2.3, and 1.4 per cent of the responses were marked incorrectly. It will be seen from the following data, however, that misscorings had a complex relationship to achievement in the 6th grade.

fact that the relationship appears suggests that cheating and the low achieve-

Rank order correlations between individual spelling gains measured on the Stanford test and frequency of cheating, misscoring, errors, and chronological age were calculated for the 6th grade experimental group and are presented in TABLE 11.

TABLE 11

RANK ORDER CORRELATIONS BETWEEN INDIVIDUAL SPELLING GAINS ON THE STANFORD TEST AND AGE AND FREQUENCIES OF CHEATING, MISSCORING, AND ERROR.

| Variable             | Rank order<br>Correlation |
|----------------------|---------------------------|
| Cheating . . . . .   | -.302                     |
| Misscoring . . . . . | -.031                     |
| Errors . . . . .     | .096                      |
| Age . . . . .        | .199                      |

With  $N = 31$ , none of the correlations are significant except that between cheating and achievement which is just significant at the .05 level for a single-tailed test. The low correlations are a result of very low percentages of misscoring and errors for all but a few individuals in the group, with the misscorings being made on 1.4 per cent of the total responses and errors on 4 per cent. The low correlation with age also represents the typical small range of ages present in one grade. The relatively high negative correlation between achievement and cheating is not the result of a great deal of cheating and a wide range of differences in cheating, since the average percentage of cheating was only .6 with a range from no cheating to 2.8 per cent. The fact that the relationship appears suggests that cheating and the low achieve-



ment may have a common cause in other factors influencing the attitudes of individuals toward the teaching machine treatment and perhaps school in general.

When the same variables are correlated with weekly spelling test scores rather than Stanford gains, interestingly different results are produced. These results are presented in TABLE 12. With an N of 30,

TABLE 12

RANK ORDER CORRELATIONS OF AVERAGE WEEKLY TEST SCORES FOR INDIVIDUALS VERSUS AGE AND FREQUENCIES OF CHEATING, MISSCORING, AND ERRORS.

| Variable             | Rank order<br>Correlation |
|----------------------|---------------------------|
| Cheating . . . . .   | -.142                     |
| Misscoring . . . . . | .631                      |
| Errors . . . . .     | -.754                     |
| Age . . . . .        | -.375                     |

the correlations are all significant except with cheating. The age correlation is significant beyond the .05 level, errors and misscoring beyond the .0005 level for a single-tailed test. The non-significant cheating correlation indicates that cheating had little influence on the learning outcomes of the teaching machine treatment, as measured by the weekly tests, which covered the specific content of the program. This finding is supported by a similar analysis of the influence of cheating upon individual lessons, to be presented in the next section. This further supports the notion that the negative relationship between cheating and Stanford test gains is of a general nature, reflecting influences outside of the teaching machine treatment. The highly significant negative correlation between

errors and test scores does not necessarily mean that more errors on the program resulted in less learning. All that can be said at this point is that the individuals who make more errors have lower weekly test scores, and that both the high number of errors and the lower scores may be caused by some common factor. The relationship to age is felt to reflect the same ceiling effects which were responsible for the negative correlation between IQ and gains noticed in the upper IQ group. The correlation between misscoring and weekly test scores is also thought to be related to this factor, and will be discussed in the section on programming factors.

Student reactions to the teaching machine treatment were obtained on the 6th grade level from a questionnaire that asked the following questions and also gave opportunity for open-ended comments.

1. How do you like to study spelling better?

\_\_\_ with machines

\_\_\_ without machines

2. When did you learn most spelling?

\_\_\_ with machines

\_\_\_ without machines

These questions were followed by others that asked for specific likes and dislikes concerning the teaching machines and suggestions for improving the technique. Results showed 17 liking machine instruction better and 12 students liking instruction without the machines. Median gain of those liking machine instruction was 1.6 grade equivalent score, for those liking other modes of instruction, 1.2. With N's of 12 and 17 and a U of 73, P is .09, a non-significant difference. Reasons volunteered for disliking machine instruction centered about three points, the machines, the materials,



and administration of the machine instruction in the classroom. Nine students complained about size, weight, and operation of the machines. Twelve said that the machines left them no more spare time, complained about having to hurry too fast, or said that other students tried to race through the work to see who could finish first. Seven students complained that the materials were unsuitable, primarily because they were not challenging enough or did not cover certain aspects of spelling. Six of the seven were in the high IQ group that produced a negative correlation between IQ scores and gains in spelling. This lends support to the explanation of the correlation given previously. Finally, the one student who said that he "disliked everything and liked nothing" about the teaching machines also claimed that he learned more with them. He also made the largest individual gain of the year, 3.8 in grade equivalent score.

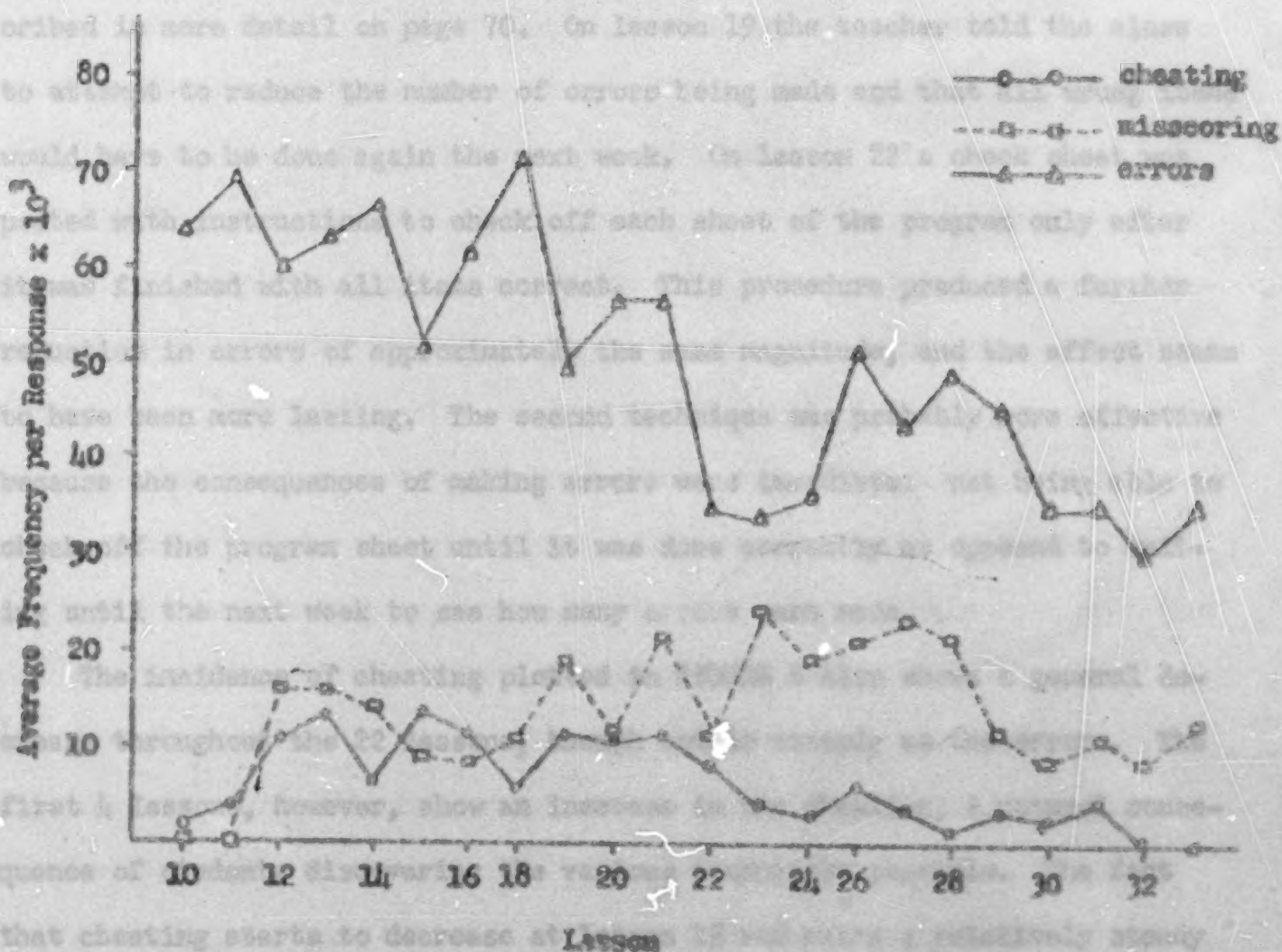
Programming factors.--The effects of cheating, misscoring, and errors upon learning outcomes can be examined from the points of view of both students and lessons. In the previous section on individual variables a strong relationship was seen between the average number of errors and misscorings made by individuals and their weekly test scores, while a significant negative relationship existed between Stanford gains and cheating. For the experimental group as a whole, then, idiosyncrasies of responding to the teaching machine program were important. Here the individual differences between lessons will be examined for the same variables, and for the number of responses per lesson, a variable which can be looked at only across lessons.

FIGURE 6 gives the average weekly incidence of cheating, misscoring, and errors for each of the 22 machine lessons, plotted as average frequencies

per response  $\times 10^3$  to adjust for differences in the number of responses required per lesson. It will be noted that the average errors per response decrease quite steadily throughout the 22 lessons but seem to approach an asymptotic value of about 30 in the last 4 lessons, an asymptotic error rate of 3 per cent. It is doubtful that this caused the drop in errors.

FIGURE 6

AVERAGE INCIDENCE OF CHEATING, MISSCORING, AND ERRORS  
FOR THE 22 MACHINE LESSONS, PLOTTED AS AVERAGE  
FREQUENCIES PER RESPONSE  $\times 10^3$





Since the "difficulty" of the spelling lessons measured by the average grade placement of the words on each lesson has a range of only 5.58 to 5.67, and the rank order correlation between difficulty and weekly test scores is a non-significant .161, it is doubtful that this caused the drop in errors. The most reasonable explanation, of course, is that students became familiar with the teaching machine procedures and the types of items used as the year went on. The greatest drops in error rate took place on lessons 19 and 22 and were the direct result of procedures initiated in the classroom and described in more detail on page 70. On lesson 19 the teacher told the class to attempt to reduce the number of errors being made and that all wrong items would have to be done again the next week. On lesson 22 a check sheet was posted with instructions to check off each sheet of the program only after it was finished with all items correct. This procedure produced a further reduction in errors of approximately the same magnitude, and the effect seems to have been more lasting. The second technique was probably more effective because the consequences of making errors were immediate: not being able to check off the program sheet until it was done correctly as opposed to waiting until the next week to see how many errors were made.

The incidence of cheating plotted in FIGURE 6 also shows a general decrease throughout the 22 lessons, though not as steeply as the errors. The first 4 lessons, however, show an increase in the cheating, a natural consequence of students discovering the various techniques possible. The fact that cheating starts to decrease at lesson 15 and makes a relatively steady decline from that point on cannot be related to any known factors in the classroom. A reasonable inference is that as the number of errors decreased

it became less necessary to cheat for the correct answer. The relationship between cheating and errors, if not the direction of causality, is shown by a rank order correlation of .630 for lessons 15 through 33. With an N of 17 the correlation is significant beyond the .01 level.

Changes in the amount of misscoring are not so regular. An initial increase similar to the cheating is followed by a 6 lesson plateau and then an increase to lesson 27. Thereafter the rate of misscoring drops to approximately the level of the initial plateau. By inspection there does not appear to be a strong relationship between misscoring and the other variables plotted in FIGURE 6, but during the rising section of the misscore plot both cheating and errors are falling. Considering together the decrease in cheating and errors, the increase in misscoring, and the previously reported decrease in weekly test scores, it is tempting to attribute all of these phenomena to a "motivational" cause with the increasing misscoring being a result of "careless indifference." One should not be caught in this kind of circular explanation, however. The problem is to explain what events in the teaching machine program and/or the classroom were responsible for the observed changes. Competing classroom activities, controlled by the teacher, could have relegated the teaching machine lessons to a spare-time and hastily done activity. Volunteered responses to the questionnaire suggest that this may have been the case: "I don't like the way we have to hurry with the machines to get all the sheets done;" "Let Miss X give us more time...;" "...and you raced through to finish them in time." Insofar as characteristics of the program itself generated the increase of misscoring it is a sign of poor programming, and the program was most certainly responsible for some misscoring. As



students became accustomed to the style of the program it was possible to respond to some items on the basis of only part of the discriminative stimuli presented, and not check the answer carefully with the program answer because it seemed quite obviously correct. For example, in item 23-3.4 spelling responses must be made to two words and a plausible but "wrong" response can be supplied in one case; Wh\_n\_v\_r we get cold and sti\_\_, there will be a bonfire. The desired responses are whenever and stiff, but still was the answer written and called correct by 6 students. If the semantic context provided by previous items had been responded to adequately, students would have realized that only stiff made the sense required by the story, but the program had enough extraneous and redundant material that students evidently learned to not read some portions of the items. Further misorderings were the result of presenting a series of items which required only one response, then giving an item requiring two responses, one of which was not responded to nor scored as being incorrect. If students had been reading the items carefully enough to get their meanings this would not have happened, but the fact that some items were not read carefully was a program flaw, not the fault of the students.

Rank order correlations of the incidence of cheating, misordering, errors, and number of responses per lesson versus average weekly posttest scores are shown in TABLE 13. The rho's are based upon an N of 22, the number of lessons involved. None of the correlations approach significance except between responses per lesson and achievement per lesson, which is just significant at the .05 level (single-tailed).

TABLE 13

RANK ORDER CORRELATIONS OF THE AVERAGE FREQUENCY PER RESPONSE OF  
CHEATING, MISSCORING, ERRORS, AND RESPONSES PER  
LESSON VS. AVERAGE TEST SCORES PER LESSON

| Variable                       | Rank order<br>Correlation |
|--------------------------------|---------------------------|
| Cheating . . . . .             | -.164                     |
| Misscoring . . . . .           | -.181                     |
| Errors . . . . .               | -.046                     |
| Responses per lesson . . . . . | .363                      |

One contributing factor to the low correlations is the small range in the variables and the high variability of weekly test scores after approximately lesson 20. It would be unwise on the basis of this data to conclude that cheating and misscoring have no influence upon the success of a teaching machine treatment. If cheating had been more frequent and had not extinguished, there is theoretical reason, discussed earlier, to suppose it would spoil the stimulus control of the program. It is also felt that the previous analysis of miscorings functioning as errors is correct, but that the data gathered to test the hypothesis was inadequate because of the low frequency of misscoring.

The highly significant correlation of .631 between misscoring by individuals and their weekly posttest scores, reported in the previous section, had its origin in the spelling program but that does not mean that making miscorings contributed to learning outcomes from use of the program. The small correlation between misscoring and weekly test scores across lessons supports this notion and suggests that the high correlation across



individuals is due to other factors. These factors probably lie within the program and the learning histories of the students. Ninety-six per cent of the miscorings consisted of calling incorrect answers correct, and were generated by characteristics of the program that resulted in the shaping-up of poor item reading and poor answer matching behavior. Careless item reading was the result of poor programming of the nature described previously in this section, where the highly predictable pattern of programming and redundancy of the text used to prompt the spelling of the correct words made it unnecessary for students to do more than just glance at an item in order to guess what word they were to spell. Under such circumstances semantically plausible but incorrect words were written in response to some items. The poor matching behavior was probably generated by the long series of correct answers most students made to the program. When the items were "easy," that is, when the students already knew most of the answers they were to supply, the confirmation provided by program answers ceased to have any real function and the students probably assumed they were correct rather than checking their answers by matching them to the program answers. Most individuals who have used a program with a low error rate have felt the effects of this "self-confirming" behavior and caught themselves almost scoring an incorrect response as correct. Misscoring represents the joint effects of this poor answer matching behavior and the poor item reading, and the plot of miscorings in FIGURE 6 represents the growth of these behaviors.

It was previously hypothesized that misscoring would be most frequent in the case of students who already knew most of the spelling they were being "taught" by the program. Such students would be more likely to assume

their responses were correct without checking with the answers given in the program because, in fact, their answers probably were correct most of the time. These students would also be the ones with the higher weekly test scores. This hypothesis is supported by the finding of a rank order correlation of .408 between the Stanford pretest scores and incidence of mis-scoring. This correlation is significant between the .01 and .05 levels for an N of 31, and supports the notion that the relationship between mis-scoring and weekly test scores for individuals is a matter of individual differences in learning history rather than an actual, positive contribution by mis-scoring to the learning outcomes. Again, it has been shown that the effects of a program depend very strongly upon the match between the subject-matter status of the students and the program content.

The final result to be discussed is the .363 correlation between responses per lesson and achievement. It has been a general finding to date in the teaching machine field that the more items in a program the better the results (Lumadaine and Olaver, 1960). If the addition of items does not expand the coverage of a program, there will certainly come a point when more items will decrease a program's efficiency and effectiveness, but evidently the 6th grade spelling program used has not reached that point. Since the added items do make a difference in learning outcomes, one can but conclude that the program is generally inadequate as it stands.

Transfer of Training.—Results of the transfer measurements for both the 4th and 6th grade groups are presented below. The 4th grade data represent the comparison of transfer effects of the two different versions of the 4th grade teaching machine program and the 6th grade data compare the teaching machine group with the teacher group.



Since individuals receiving experimental (transfer) and control versions of the program were equally distributed in the two 4th grade classrooms, teacher influence and other factors inherent in the different classrooms were well controlled. The experimental and control groups were also evenly matched on other relevant variables. TABLE 4, on page 65, reports the matching on IQ, mean achievement test scores (less spelling), and gains (less spelling). With P's of .32, .32, and .29 respectively, there are no significant differences on these variables. Matching of the groups on further variables is reported below in TABLE 4.

TABLE 4  
COMPARISON OF 4TH GRADE TRANSFER (E) AND NON-TRANSFER (C) GROUPS  
ON MATCHING VARIABLES

| Variable            | Median Scores |     | Diff. | n/n   | P*  |
|---------------------|---------------|-----|-------|-------|-----|
|                     | E             | C   |       |       |     |
| Spelling Pretest    | 518           | 531 | 13    | 30/33 | .17 |
| Spelling Gains      | 42            | 44  | 2     | 30/33 | .31 |
| California Spelling | 6.1           | 6.1 | 0     | 30/33 | .23 |
| Cal. Spelling Gains | .3            | .2  | .1    | 30/33 | .33 |
| Handwriting Scale   | 5             | 6   | 1     | 30/33 | .44 |
| Word Matching Test  | 53            | 53  | 0     | 29/32 | .47 |

\*Mann-Whitney U test, single-tailed.

There are no significant differences between the groups on any of these variables, either, indicating a very close match between the two groups. The gains of 42 and 44 on the spelling test specific to the words taught

are highly significant in themselves, even though the differences between them is not. For both the experimental and control groups every subject made a gain from pretest to posttest. The gains were tested for significance with the Wilcoxon matched-pairs test, using each student's pair of test scores. With a T of zero in each group and N's of 30 and 33, the differences are significant beyond the .00003 level, indicating that the treatments produced definite outcomes.

Transfer of training produced by the teaching machine treatment was measured on a test patterned closely after the regular spelling pre-posttest but containing words that were not taught. All of the words, however, contained sequences of sounds and letters which had been taught. The transfer test as well as the regular spelling test was scored on a letter rather than word basis, with the transfer test being scored only for the sequences of letters for which transfer could have taken place. The 4th grade spelling test and transfer test may be seen in APPENDIX C.

Results of the differential treatment by the two versions of the 4th grade program show a transfer score of 118 for the transfer group and 123 for the non-transfer group. Differences between the groups were tested by the Mann-Whitney U test. With  $n_1 = 30$  and  $n_2 = 33$ , the P of the difference is .18, not significant. An analysis of covariance was performed on the same data, using the regular spelling pretest as the control variable because it had the highest correlation with the transfer test scores ( $r_{ho} = .691$ ) and showed the greatest difference between the groups. The F obtained had the value of 1.00, which is not significant.



The 6th grade transfer data were obtained from 8 comparable written papers in the experimental and control classrooms, representing an average of 771 words of written copy per student. Four of the papers were rough drafts and the other four finished works, in each group. Misspelled words were scored on both a word and letter basis and averaged for each individual in the two groups. These averages were then converted into percentage errors of total words written per individual in order to adjust for differences in writing output and thus chance for misspelling. The results are presented in TABLE 15.

TABLE 15

COMPARISON OF AVERAGE PERCENTAGE SPELLING ERRORS IN WRITTEN WORK  
FOR 6TH GRADE TEACHING MACHINE AND TEACHER GROUPS

| Measure               | Group |     | Diff. | $n_1/n_2$ | P*    |
|-----------------------|-------|-----|-------|-----------|-------|
|                       | E     | C   |       |           |       |
| Average Word Errors   | 3.8   | 2.1 | 1.7   | 26/31     | .0007 |
| Average Letter Errors | 3.8   | 5.7 | 1.9   | 26/31     | .41   |

Mann-Whitney U test, single-tailed.

The teaching machine group made significantly more word errors in spelling than the teacher taught group. Although the magnitude of the difference is not large it is consistently in favor of the teacher group. On the other hand, the difference in the letter errors is not significant because of the higher variability in the data.

The negative results in the 6th grade are probably a result of several factors. The major factor is that the measure of transfer is very inadequate

because the sample of writing is small, being based upon the misspelling of 0 to 13 words per individual. Less than 1 per cent of the misspelled words were actually taught in the spelling program. The transfer measure more likely tapped long-term differences in the students or differences in procedures of the two classrooms and teachers. The teaching machine program was expected to increase transfer to writing by breaking up the immediate memory effects from repetitious study of one word and by having students write the words into meaningful sentence contexts. Evidence presented in the section on programming factors has shown that the students, in some cases, did not respond to the teaching items in the way intended. Rather than reading the sentence contexts, they were able to respond on the basis of remembering what words were to be spelled or by simply looking at the letters of the word which had not been deleted. This eliminated the basis for the expected transfer effects from the program. The other programming factor opposing effective transfer was the content of the program, which was not constructed to reveal effectively the sound-letter structure of English spelling. The control group was learning the same content, of course, so this factor does not explain the differences between the groups.

The failure of the 4th grade experimental program to produce increased transfer is attributable to the nature of the program content and the type of behavior the program attempted to teach. Program content was centered upon words for which there are not good spelling rules because the sound-letter relationships involved are irregular, and the program was designed to extinguish student use of sound cues for the spelling of these words and



others in the same sound and letter categories. Apparently 41 items were not enough to change the long-term procedure of "sounding-out" a word, even for the limited set of words used.

The negative results from these attempts to teach and measure transfer in the 4th and 6th grades should not be interpreted as showing that teaching machine instruction does not produce transfer. The use of teaching machines has only a second order influence upon whether or not transfer takes place. The prime determiners of transfer in spelling are the content of the teaching programs used and the procedures adopted by the teacher to encourage accurate spelling in written work. It might very well be that classroom adoption of teaching machine instruction in spelling would result in the teacher paying less attention to spelling because he was no longer personally responsible for it, but this is not a fault of the teaching machines any more than it is a fault of normal spelling texts.

Learning outcomes and efficiency--Over-all learning outcomes from the teaching machine treatment were generally comparable to those from the usual classroom techniques employed in the experimental school, with some suggestion of superiority for the machine teaching. When efficiency of the teaching techniques was taken into account, however, the machine treatment resulted in a time advantage of approximately 50% for the machine group. With superior students free to learn at their own pace, the gain could be even greater, especially if large sequences of the exercises were prepared so that there was no subject-matter ceiling on the program. The greatest gains in spelling were

## SUMMARY AND CONCLUSIONS

The major findings of this research are summarized below, and suggestions are made for the guidance of future research in programming and spelling.

Teaching Machines for Programming Research. The teaching machines designed for this research proved to be satisfactory and durable in operation, but left much to be desired in the control of cheating, security of the teaching materials, and general convenience. For research purposes, a form of teaching machine that allows simple preparation of materials and ease in revision is desirable and the design of such a device is suggested. It would retain the feature of collecting student responses directly on the programmed materials as an aid to item analysis, yet accommodate many sheets of material at one loading. More complex machinery to make possible the presentation of auditory stimuli and the gathering and evaluation of student responses from a keyboard would contribute to future research into the nature of spelling and other aspects of language behavior.

Learning outcomes and efficiency. Over-all learning outcomes from the teaching machine treatments were generally comparable to those from the usual classroom techniques employed in the experimental school, with some suggestion of superiority for the machine teaching. When efficiency of the teaching techniques was taken into account, however, the machine treatment resulted in a time advantage of approximately 3:1 for the entire group. With superior students free to learn at their own rates, the gain would be even greater, especially if large sequences of the curriculum were prepared so that there was no subject-matter ceiling on the program. The greatest gains in spelling were



recorded by students in the lower IQ half of the teaching machine groups, the least gains by individuals in the comparable teacher groups. Since the subject-matter level of the program matched both lower groups equally, the teaching machine treatment was responsible for the greater gains. In the upper IQ groups the machine treatment produced a non-significant advantage. The smaller gains in the upper groups were primarily a result of poor matching between subject-matter in the program and the subject-matter level of the students. In addition to limiting the possible gains this poor matching helped to generate modes of response to the teaching machine program that will be discussed below. Contrary to expectations, IQ correlations with spelling gains were heightened by the teaching machine treatment when the subject-matter matching was good. It should not be concluded that teaching machine procedures are most suitable for students with low abilities. Given appropriate program content without a low ceiling, the superior students should produce even greater gains.

Over the entire period of the experiment a drop in achievement was observed, but it was no greater on the teaching machine lessons than on teacher-taught lessons, and the trend was toward a smaller loss on the machine lessons. The drop in achievement appears to have been a result of general classroom operations and administration of the programs rather than an inherent characteristic of the teaching machine instruction.

It can be speculated that with suitable programs covering a large span of curriculum, with convenient teaching machines, and with adequate procedures for administering and supervising use of the programs the gains reported here could be at least doubled, on the average.

The measurement of spelling achievement.—Several different forms of achievement measures were used in this research. Standardized achievement tests with a general coverage of the curriculum, specially constructed tests specific to the words taught, and a transfer test containing untaught words chosen from the same sound-letter categories as words in the program were all utilized, but none of the measures were entirely suitable. The following observations are made with a view toward designing more suitable measures for future work.

Standardized achievement tests in spelling generally have only a small overlap with the curriculum of one year's spelling instruction. Typically, less than 10 per cent of the year's spelling words will appear on such a test. The fact that such tests can give a valid measurement of learning outcomes on the basis of so few items suggests that other words on the tests contain internal word elements important within the curriculum. If this is the case, the usual method of scoring a spelling test by marking entire words correct or incorrect throws away important information. The procedure followed here whenever possible was to score on the basis of single letters, with the test score representing the total number of correct letters in their proper positions in the words. Although this is an improvement on the usual procedure because information is not discarded and the effective number of responses counted is increased, it is not the ideal procedure because individual letters are not the basic structural elements of spelling. Ideally, the scoring of a spelling test should be based upon an adequate analysis of the sound-letter relationships in English spelling, just as much as the curriculum should. In the absence of an adequate analysis of the structure of English spelling a letter scoring



procedure is the best approximation and recommended for future investigations. The question of comparative validities of the word and letter scoring methods is a matter for future research. The same recommendations hold for transfer tests, but it is even more important that such instruments be based upon an adequate analysis of English spelling.

English spelling and transfer.—The design of adequate teaching materials and instruments to assess learning outcomes in spelling depends upon knowledge of the structure of English phonics, the sound-letter relationships that make it possible to predict the letter representations of particular sounds. These relationships are complex and often not helpful in English when only the first-order sound to letter correspondences are considered, but if the relationships between sequences of sounds and letters are considered further regularities in spelling emerge. That is, if the probabilities of particular spellings of a sound are considered in terms of preceding and following sounds, the chances of guessing the correct spelling are increased. A sample analysis of the spellings of the sound /iy/ proved fruitful in this respect and suggests that further analysis be undertaken to assess the generality of the findings. The design of spelling materials to produce positive transfer depends upon the success of such analysis.

Programming factors.—A number of characteristics of the 6th grade program and factors relating to its use were investigated. The average incidence of student errors, cheating, and misscoring was 4, .6, and 1.4 per cent respectively of the total responses, indicating that the program successfully met the criterion of low error rate assumed to be a characteristic of good programs. Cheating did occur, and its incidence rose from an initial zero rate to a

maximum on the fourth lesson and thereafter decreased to zero again on the next to last lesson. The initial rise can be attributed to students' discovery of the various cheating procedures possible and the eventual drop to the correlated drop in errors which made cheating unnecessary. The sections of the program containing the greater number of items per word produced significantly greater learning outcomes, indicating that despite the apparent success of the program as a teaching instrument, revisions are in order. A strong, significant, negative relationship was found between the number of errors made by individuals and their weekly test scores and a not quite significant relationship between errors per lesson and the same test scores. With a greater range in weekly test scores than was obtained, the correlation between errors per lesson and weekly test scores probably would have been significant, indicating the detrimental effects of error responses upon learning from the program. On the other hand, the fact that the students who made the greater number of errors also received the lower weekly test scores does not necessarily indicate a causal relationship. Both the errors and poor test showing could have had a common cause in work habits of long standing.

Although misscoring of responses—primarily calling them correct when they were incorrect—took place on only 1.4 per cent of all responses, the discovery of how it developed has important implications for the writing of future teaching materials. There were two components of the misscoring: inadequate reading of the teaching items, and students not matching their own responses with the answers given in the program. The poor item reading was generated by redundant teaching items that made it unnecessary for students to do much more than glance at the items in order to determine what words they



were being requested to spell. On occasion, however, the semantic context provided by the sentences in which the spelling words were to be written provided important cues which were not noticed because of the poor reading. This led to the spelling of the wrong words. The miscoverings took place on items of this sort where the students did not bother to match their responses with the program answers. The poor answer matching was also generated by the program. Since the program produced a very low student error rate, the answer matching response served no function most of the time and tended to extinguish. This was expected most often in the records of students who already knew most of the spelling content of the program, and was supported by a significant positive correlation between pretest scores and incidence of miscovering.

If programs are to be written with very low student error rates, there is probably no way of eliminating occasional non-matching so long as students judge their own responses. Insofar as teaching machine programs are not perfect and students do make errors, letting the errors go by without correction may be detrimental to learning outcomes. This problem can be solved by having the machine itself judge the responses and require a correction before going on to the next item. In addition to eliminating the non-matching problem, this procedure does not give away the answer by showing it to the student, but requires another attempt under stimulus control of the teaching material contained in the item.

Errors made in response to program items were of two major types: where spelling of the wrong word was attempted, and where the right word was spelled incorrectly. Errors of the first type were a result of inadequate semantic

prompting, while errors of the second type were a result of inadequate phonics prompting. Both types of errors should be treated by re-programming. Semantic errors could be eliminated by reducing the between and within item redundancy so that students are forced to read rather than glance at the items. Items that require a semantic discrimination within themselves would be especially useful in this respect. For example: "We wash our hands with so\_p and drink hot so\_p." The elimination of phonics errors depends upon a thorough reorganization of the spelling teaching sequences, based upon better information about English phonics. In the cases where sound-letter relationships are not regular enough to offer any assistance, use should be made of any semantic, letter pattern, or letter position contingencies which are available. For example, teaching the spelling of here and there, and hair and their in the same context would make it possible to point up the semantic and spelling parallels between the words.

Concluding remarks.—This research has considered the teaching of spelling as a phenomenon more or less isolated from the other areas of language arts instruction that probably contribute to spelling performance. Various aspects of reading instruction and the creative writing experiences of students certainly have some influences upon the nature of spelling behavior. The phonics instruction of early reading may contribute both positively and negatively toward spelling, and spelling on written work must depend to some extent upon teacher imposed standards of accuracy. Nevertheless, within the limited framework of this research an interesting new teaching methodology has been tried with provocative results, and some steps have been taken toward a better understanding of spelling as a subject-matter.

①  
A  
④



# BIBLIOGRAPHY

- Armons, R. B. Knowledge of performance: Survey of literature, some possible applications, and suggested experimentation. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, Aero Medical Laboratory, February, 1954. (WADC Technical Report 54-14.)
- Arps, O. F. A preliminary report on work with knowledge versus work without knowledge of results. Psychol. Rev., 1917, 24, 449-455.
- Bagley, W. C. The educative process. New York: Macmillan, 1906.
- Berlyne, D. E. Conflict, arousal, and curiosity. New York: McGraw-Hill, 1960.
- Bijou, S. W. Methodology for an experimental analysis of child behavior. Psychol. Reports, 1957, 3, 243-250.
- Ferster, C. B. and Sapon, S. M. An application of recent developments in psychology to the teaching of German. Harv. educ. Rev., 1958, 28, 58-69.
- Ferster, C. B. and Skinner, B. F. Schedules of reinforcement. New York: Appleton-Century-Crofts, 1957.
- Fitzgerald, J. A. A basic life spelling vocabulary. Milwaukee: Bruce, 1951.
- Gates, A. I. New methods in primary reading. New York: Bureau of Publications, Teachers College, Columbia University, 1928.
- Gates, A. I. Generalization and transfer in spelling. New York: Bureau of Publications, Teachers College, Columbia University, 1935.
- Gates, A. I. Spelling difficulties in 3876 words. New York: Bureau of Publications, Teachers College, Columbia University, 1937.
- Harris, P. R. and Moore, J. T. Spelling—from spoken word to written symbol. Elem. Sch. J., 1953, 53, 329-337.
- Harlow, H. F. Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. J. Comp. Physiol. Psychol., 1950, 43, 289-294.
- Henry, Jules, Working paper on creativity. Harv. educ. Rev., 1957, 27, 148-155.
- Holland, J. O. Human vigilance. Science, 1958, 128, 61-63.

- Holland, J. G. Teaching machines: an application of principles from the laboratory. Proceedings of the Educational Testing Service Invitational Conference, October 1959, on "The Impact of Testing on the Educational Process." Educational Testing Service, 1960.
- Horn, Ernest. Spelling. Encyclopedia of educational research (rev. ed.), 1950, 1347-1354.
- Horn, Ernest. Phonetics and spelling. Educ. sch. J., 1957, 57, 421-432.
- Horn, Ernest. Spelling. Encyclopedia of educational research (3rd ed.), 1960, 1337-1354.
- Judd, C. H. Practice without knowledge of results. Psychol. Rev. Monogr. Suppl., 1905-6, 7, 185-198.
- Lumsdaine, A. A. and Glaser, Robert. Teaching machines and programmed learning. Washington: National Education Association, 1960.
- Mallan, Thert. Teaching and educational inventions. J. exp. Educ., 1935, 4, 291-300.
- Mendenhall, J. E. An analysis of spelling errors. New York: Bureau of Publications, Teachers College, Columbia University, 1930.
- Miller, G. A. Language and communication. New York: McGraw-Hill, 1951.
- Moore, J. T. Phonetic elements appearing in a 3000 word spelling vocabulary. Unpublished Doctoral Dissertation, School of Education, Stanford University, 1951.
- Popp, Helen M. and Porter, D. Programming verbal skills for primary grades. Audio visual communic. Rev., 1960, 8, 165-175.
- Porter, D. A critical review of a portion of the literature on teaching devices. Harv. educ. Rev., 1957, 27, 126-147.
- Pressey, S. L. A simple apparatus which gives tests and scores--and teaches. School and Society, 1926, 23, 373-376.
- Pressey, S. L. A third and fourth contribution toward the coming 'industrial revolution' in education. School and Society, 1932, 36, 668-672.
- Rineland, H. D. A basic vocabulary of elementary school children. New York: Macmillan, 1945.
- Rothkopf, E. Z. Heuristic discussions of psychological bases for the conduct of training by automatic devices: I. A functional diagram and an approach to research. Unpublished Staff Memorandum, Maintenance Laboratory, AFPTC, Lowry Air Force Base, Colorado. April, 1957.



- Ruediger, W. C. The indirect improvement of mental function through ideals. Educ. Rev., 1908, 26, 364-371.
- Siegel, Sidney. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.
- Skinner, B. F. The science of learning and the art of teaching. Harv. educ. Rev., 1954, 24, 86-97.
- Skinner, B. F. A case history in scientific method. American Psychologist, 1956, 11, 221-233.
- Skinner, B. F. Teaching machines. Science, 1958, 126, 960-977.
- Svensen, Esther J. and Caldwell, C. G. Spelling in childrens' letters. Elem. sch. J., 1948, 49, 224-235.
- Thorndike, E. L. Handwriting. New York: Teachers College, Columbia University, 1917.
- Thorndike, E. L. and Lorge, I. The teacher's word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.
- Wart, J. E., Heidt, C. O. and Abson, J. S. Statistical methods in educational and psychological research. New York: Appleton-Century-Crofts, 1954.
- Wolff, Sister Maria R. A study of spelling errors with implications concerning pertinent teaching methods. Elem. sch. J., 1952, 53, 458-466.

# High Grade Program

Sample of item types.

Read and answer the question.

## APPENDIX

Item

THIRD GRADE STUDENT

Underline these words: receive, tires, sight.

Dear Sir: I am writing again to ask when I can expect to receive the

**APPENDIX A: Sample lesson materials from the 6th and 4th grade Programs.**

**APPENDIX B: Spelling words taught in the 6th, 4th, and 2nd grade Programs.**

**APPENDIX C: Fourth grade pre-posttest and transfer test.**

He wanted his money returned. The tires had not been received.

The tires had not been received.

Circle the words with the letters r i in them.

receive Sept. French article sight view

receive

Sept.

sight

Circle the word that means to get something that is given or sent to you.

Sept. send receiver tires receive perhaps

receive

Write the missing letters. They are all the same.

When can I expect to r \_ \_ \_ \_ the bicycle tires I ordered?

receive

Write the missing letters. They are both the same.

I ordered the r \_ \_ \_ \_ on the 4 \_ \_ \_ \_ of April.

tires

sight

Write the missing letters. Both pairs are the same.

I did not see \_ \_ \_ \_ on the tires even by the \_ \_ \_ \_ path of my.

receive

sight

Write the missing letters.

When can I expect to r \_ \_ \_ \_ the "offender's" tires?

-117-

receive



Sixth Grade Program

Sample of item types.

Read and answer the questions:

Name \_\_\_\_\_

THOSE BICYCLE TIRES

Underline these words: receive, tires, eighth

Dear Sir: I am writing again to ask when I can expect to receive the bicycle tires I ordered on the eighth of April.

tires

eighth

receive

John wrote the letter because: \_\_\_\_\_ the tires were not good.  
\_\_\_\_\_ he wanted his money returned. \_\_\_\_\_ the tires had not been received

the tires had not been received

Circle the words with the letters e i in them:

receive depot freight article eighth view

receive

freight

eighth

Circle the word that means to get something that is given or sent to you:

depot send receiver tires receive purchase

receive

Write the missing letters. They are all the same.

When can I expect to r \_ c \_ iv \_ the bicycle tires I ordered?

receive

Write the missing letters. They are both the same.

I ordered the t \_ res on the e \_ ghth of April.

tires

eighth

Write the missing letters. Both pair are the same.

I did not rec \_ \_ ve the tires even by the \_ \_ ghth of May.

receive

eighth

Write the missing letters.

When can I expect to r \_ c \_ \_ ve the "A&W's" tires?

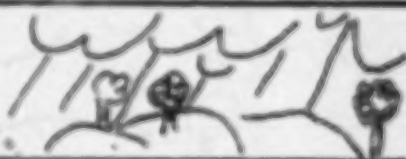
receive

NAME \_\_\_\_\_

6.31S-8

Write the missing letters; they are all the same:

Ragged clouds were sw\_\_\_\_\_ping the south\_\_\_\_\_rn sky.



sweeping

southern

Write the missing letters:

Without thinking of an umbre\_\_\_\_\_a, I set out.



umbrella

Write the missing letters:

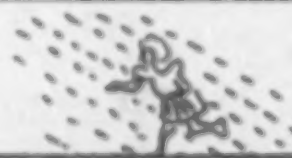
Halfway to the store thu\_\_\_\_\_er rolled.

BOOM!

thunder

Write the missing letters:

Then rain came down in a s\_\_\_\_\_ea\_\_\_\_\_y pour.



steady

Write the missing letters:

Brushing, moving quickly, s\_\_\_\_\_ee\_\_\_\_\_ing.



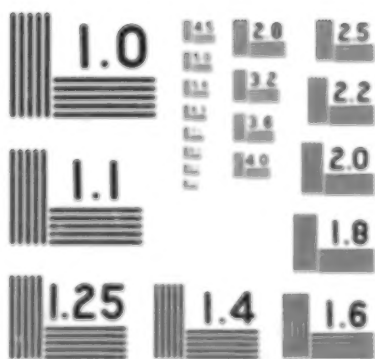
sweeping



3 OF 3

ED

002772



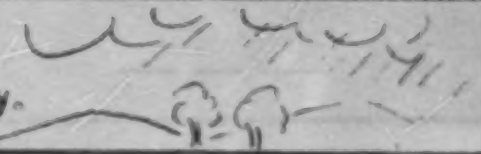
MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963

NAME \_\_\_\_\_

6.31S-9

Write the missing letters:

Ragged clouds were sweeping the s \_\_\_\_\_ th \_\_\_\_\_ sky.



southern

Write the missing letters:

Without thinking of an um \_\_\_\_\_ ll \_\_\_\_\_, I set out.



umbrella

Write the missing letters:

Halfway to the store t \_\_\_\_\_ de \_\_\_\_\_ rolled.

BOOM BOOM !

thunder

Write the missing letters:

Then rain came down in a st \_\_\_\_\_ d \_\_\_\_\_ pour.



steady

Write the missing letters:

Brushing, moving rapidly, \_\_\_\_\_ ee \_\_\_\_\_

sweeping



Fourth Grade Program

NAME \_\_\_\_\_

Selected items from the alphabetical review.

What are the first 3 letters of the alphabet?

ab c d e f g h i j k l m n o p q r s t u v w x y z

a b c

Write the missing letters.

ab c d \_ \_ h i j k l m n o p q r s t u v w x y z

e f g

Write the missing letters.

ab c \_ e f g \_ i j k \_ m n o \_ q r s \_ u v w x y z

d h l p v

Write the missing letters.

ab c d e g h i j l m n o q r s t u x y z

f k p v

Fourth Grade Program

NAME \_\_\_\_\_

Sheet 27, experimental version of program.

You should cross the street carefully.

four

pour

or

cook

took

?

Does should look more like -----

soup

look

should looks more like -----

four

pour

soup

Write the 2 letters that are the same in all these words: should

pound

ground

\_\_\_\_

The two letters that are the same are:

o u

Write the missing letters.

s h o u \_\_\_\_ d

The missing letter in should is:

l

Write the missing letters.

\_\_\_\_ o u l d

The missing letters in should are:

s h

Write the missing word

You s \_\_\_\_ cross the street carefully.

\_\_\_\_\_

The missing word is spelled:

s h o u l d



Fourth Grade Program

NAME \_\_\_\_\_

Sheet 27, control version of program.

You should cross the street carefully.

Write the third and fourth letters of should.

The third and fourth letters in should are: o u

Write the missing letters.

s h \_ \_ l \_ \_

The missing letters in should are: o u d

Write the missing letters.

s h o u \_ d

The missing letters in should are : l

Write the missing letters.

\_ \_ \_ \_ \_ ould

The missing letters in should are: s h

Write the missing word.

You s \_ \_ \_ cross the street carefully.

The missing word is spelled: should

Fourth Grade Program

NAME \_\_\_\_\_

Sheet 78, review of previously taught words.

The car skidded on the slippery s\_r\_f\_c\_e.

Write the missing letters of the word.

s \_ \_ r f \_ \_ c e

The missing letters in surface are:

u a

We saw clowns at the c\_r\_c\_s.

Write the missing letters of the word.

c \_ \_ r c \_ \_ s

The missing letters in circus are:

i u

Please be very qu\_ \_ t.

Write the missing letters of the word.

q u \_ \_ \_ t

The missing letters in quiet are:

i e

Write the missing word.

The car skidded on the slippery s \_ \_ \_ \_ .

s \_ \_ \_ \_

The missing word is spelled:

s u r f a c e

Write the missing word.

Please be very q \_ \_ \_ .

q \_ \_ \_ \_

The missing word is spelled:

q u i e t



# APPENDIX B. SPELLING WORDS TAUGHT

## Words in Second Grade Program

| Lesson No.                                   | Lesson No.                                       | Lesson No.                                                 | Lesson No.                                                          | Lesson No.                                                               |
|----------------------------------------------|--------------------------------------------------|------------------------------------------------------------|---------------------------------------------------------------------|--------------------------------------------------------------------------|
| 1) to<br>I<br>a<br>do                        | 10) an<br>come<br>all<br>out<br>with<br>there    | 17) said<br>him<br>be<br>horse<br>this<br>has              | 23) about<br>once<br>night<br>ride<br>green<br>fish<br>box          | 29) dear<br>apples<br>thank<br>father<br>apple<br>name<br>if<br>bring    |
| 2) is<br>he<br>me<br>on                      | 11) some<br>went<br>man<br>up<br>make<br>day     | 18) took<br>put<br>jump<br>made<br>pretty<br>take<br>very  | 24) bird<br>where<br>find<br>why<br>things<br>birds<br>around       | 30) winter<br>cold<br>morning<br>work<br>fly<br>blue<br>cats<br>children |
| 3) it<br>the<br>see<br>go<br>of              | 12) they<br>was<br>get<br>run<br>your<br>saw     | 19) comes<br>ran<br>after<br>gave<br>many<br>give<br>baby  | 25) lives<br>live<br>train<br>by<br>who<br>bed<br>help              | 31) letter<br>every<br>think<br>boat<br>street<br>yard<br>hill<br>sat    |
| 4) too<br>and<br>in<br>at<br>she             | 13) mother<br>play<br>have<br>home<br>time<br>no | 20) from<br>plays<br>away<br>over<br>down<br>other<br>back | 26) here<br>new<br>black<br>bears<br>were<br>bear<br>us<br>soon     | 32) story<br>best<br>two<br>dolls<br>pig<br>head<br>store<br>sleep       |
| 5) you<br>my<br>girl<br>her<br>can           | 14) his<br>do<br>what<br>tree<br>eat<br>how      | 21) into<br>look<br>brown<br>yes<br>fast<br>eats<br>wish   | 27) happy<br>long<br>runs<br>four<br>our<br>water<br>snow<br>now    | 33) last<br>toy<br>books<br>toys<br>five<br>book<br>pony<br>car          |
| 6) like<br>likes<br>are<br>will<br>one       | 15) school<br>fun<br>that<br>white<br>old<br>pet | 22) yellow<br>three<br>dogs<br>got<br>hen<br>them<br>when  | 28) door<br>rabbit<br>milk<br>eggs<br>kitten<br>under<br>just<br>as | 34) room<br>hat<br>town<br>coat<br>fat<br>sled<br>told<br>men            |
| 7) for<br>we<br>little<br>going<br>did       | 16) ball<br>then<br>boys<br>cat<br>dog<br>girls  |                                                            |                                                                     |                                                                          |
| 8) big<br>good<br>dog<br>want<br>boy         |                                                  |                                                            |                                                                     |                                                                          |
| 9) had<br>not<br>but<br>red<br>house<br>came |                                                  |                                                            |                                                                     |                                                                          |

Words in Fourth Grade Program\*

|             |             |             |
|-------------|-------------|-------------|
| full        | + through   | usually     |
| hose        | + pour      | together    |
| banana      | + surface   | succeeded   |
| paint       | + fountain  | + improve   |
| crowd       | + equals    | + which     |
| could       | + crowded   | + necessary |
| beautiful   | + ought     | + breakfast |
| throat      | circus      | + Tuesday   |
| ground      | woman       | + although  |
| wolf        | + famous    | + height    |
| pleasant    | chalk       | + sugar     |
| + wooden    | + because   | second      |
| + bushes    | broad       | + daughter  |
| direction   | + purchase  | thousand    |
| + should    | + quiet     | + excuse    |
| minute      | + reason    | October     |
| + flood     | + factories | soap, soup  |
| + witches   | + trouble   | fourteen    |
| + author    | + suppose   | become      |
| tomorrow    | dried       | lion        |
| + countries | + against   | raise       |
| + tooth     | exhibit     | village     |
| awful       | February    | + pound     |

\* Words in this program were not arranged by lessons, but are listed in the order in which taught.

+ Words marked with a plus (+) were treated differentially in experimental and control versions of the 4th grade program.



# Words in Sixth Grade Program

## Lesson No.

1--not prepared

2--not prepared

3--not prepared

4--not prepared

5--not prepared

6--not prepared

7--not prepared

8--not prepared

9--not prepared

10--not prepared

stack  
chapter  
foolish  
burst  
sister's  
settle  
headache  
ruin  
regular  
upset  
outline  
travel  
useless  
whatever  
important  
whisper

## Lesson No.

11--arrived

eastern  
hearing  
crash  
shock  
central  
flight  
rattle  
pilot  
soldiers  
airport  
passengers  
section  
landing  
final  
alarmed  
spin

12--damaged

model  
cleaner  
preacher  
rude  
married  
double  
contain  
fairly  
correct  
spoiled  
review  
ruler  
bold  
bound  
gained  
runner

13--ski

length  
chin  
setting  
attention  
reasons  
cement  
pleased  
special  
stunts  
seventeenth  
yelled  
nicely  
treat  
stumbled  
lip  
cheerful

## Lesson No.

14--favor

upstairs  
object  
canning  
chimney  
entertain  
keen  
china  
kettle  
ghost  
visitor  
swept  
closet  
rifle  
offered  
noticed  
fireplace

15--forenoon

bushel  
chewing  
we're  
one's  
trim  
ladder  
peanut  
current  
starving  
youngest  
cord  
funniest  
theirs  
member  
direct  
manner

16--fountain

thirsty  
weave  
suffer  
improve  
message  
drain  
steel  
weaving  
drug  
struck  
shown  
odd  
swell  
hospital  
lack  
magic

Words in Sixth Grade Program (Continued)

Lesson No.

17--not prepared

18--noisy  
whom  
violin  
remarks  
we've  
period  
squeeze  
recite  
worry  
quietly  
file  
closely  
singers  
excited  
provide  
rapid  
although

19--examinations

stack  
chapter  
foolish  
burst  
sister's  
settle  
headache  
ruin  
regular  
upset  
outline  
towel  
useless  
whatever  
important  
whisper

Lesson No.

20--famous

scale  
weary  
desert  
single  
castle  
dwarf  
labor  
raising  
harder  
modern  
amount  
split  
slaves  
mighty  
tons  
haul

21--ditch

garage  
lend  
remove  
accept  
easily  
heels  
driver  
gentleman  
tune  
traveling  
pity  
greeted  
pocketbook  
speed  
automobiles  
growth

22--what's

bacon  
bowl  
apart  
closing  
toast  
tickled  
stirred  
shallow  
slice  
let's  
prepared  
dough  
popped  
fried  
pantry  
carefully

Lesson No.

23--remind

include  
zero  
stiff  
knitting  
skating  
twenty-four  
bonfire  
lap  
sweater  
fourteenth  
mild  
ourselves  
saucer  
gang  
sports  
whenever

24--nation

artist  
woolen  
quilt  
simple  
lid  
colt  
socks  
eagle  
idle  
chest  
clover  
copper  
understood  
pattern  
slippers  
screen

25--not prepared



# Words in Sixth Grade Program (Continued)

## Lesson No.

26--desire

aid  
scratch  
stump  
scarce  
postman  
owner  
select  
enemy  
driven  
route  
jet  
average  
sparrows  
delighted  
chosen  
weeds

27--canal

forever  
climate  
swamps  
spade  
connect  
owe  
sickness  
population  
wipe  
greater  
tough  
fever  
united  
honor  
earlier  
germ

28--receive

depot  
mixed  
explained  
deliver  
tires  
weren't  
lying  
bocklet  
shipping  
eighth  
freight  
delay  
supply  
mentioned  
article  
shouldn't

## Lesson No.

29--range

strap  
action  
prevent  
tame  
training  
saddle  
wrist  
rid  
willing  
however  
stepped  
mule  
tight  
intend  
forward  
contest

30--writer

errand  
rank  
sword  
scarf  
someday  
subject  
gown  
horseback  
winner  
tale  
likely  
waist  
stolen  
secret  
film  
spear

31--ragged

knit  
thunder  
shadow  
space  
sweeping  
general  
steady  
frightened  
lasted  
southern  
carries  
pour  
licked  
yarn  
umbrella  
soaked

## Lesson No.

32--teased

port  
rod  
healthy  
dip  
sailor  
harbor  
exchange  
tanned  
scraps  
hired  
pier  
greetings  
view  
homesick  
timber  
stopping

33--hopping

spinach  
products  
rake  
believing  
willow  
unable  
appear  
letting  
burnt  
stranger  
lemon  
speaking  
you've  
nickel  
wax  
husband

34--not prepared

## APPENDIX C

### Fourth Grade Pre-posttest

sample:  
tooth  
bushes  
breakfast  
pour  
crowded  
Tuesday  
flood  
although  
awful  
February  
quiet  
should  
broad  
circus  
countries  
woman  
famous  
would  
witches  
wooden  
through  
dried  
tomorrow  
trouble  
reason  
ought  
against  
minute  
fountain  
because  
throat  
together  
author  
equals  
suppose  
direction  
usually  
pleasant  
improve  
surface  
factories  
purchase  
succeeded  
necessary  
exhibit

#### STORE

The dentist pulled his tooth.  
He hid behind the bushes.  
We eat grapefruit for breakfast.  
Pour some milk into this glass.  
The room was crowded.  
Tuesday comes after Monday.  
Too much rain may cause a flood.  
Although it was cold he wore no coat.  
It made an awful crash when he dropped the dishes.  
February comes after January.  
It is quiet at night.  
You should come to school on time.  
The Mississippi river is very broad.  
The circus is coming soon.  
There are many countries in Europe.  
The name of that woman is Mrs. Jones.  
Boston is a famous city.  
Would you come with me?  
Halloween is a time for witches.  
My screwdriver has a wooden handle.  
We drove through Cambridge.  
The wet shoes dried by the fire.  
We are going to the movies tomorrow.  
We had trouble with the car.  
Have you a good reason for leaving?  
You ought to know better!  
He threw the ball against the wall.  
Just a minute!  
There is a fountain in the park.  
I like winter because of the snow.  
His throat was dry.  
He and I play together.  
She is the author of that book.  
Two plus two equals four.  
Suppose it snows in the summer; what would you do?  
My house is in that direction.  
We usually go away in the summer time.  
The trip was very pleasant.  
Try to improve your spelling.  
The surface of the ocean was calm.  
Furniture is made in factories.  
They are going to purchase a brand new car.  
After trying three times, he finally succeeded.  
Food, air, and water are necessary to keep us alive.  
We saw an exhibit of animal pictures.



### Fourth Grade Transfer Test

sample:

tools  
pushed  
weather  
soul  
powder  
blood  
shoulder  
lawyer  
January  
diet  
could  
board  
circle  
cities  
nervous  
smooth  
itches  
woolen  
group  
tried  
borrow  
double  
season  
bought  
captain  
lettuce  
mountain  
taught  
float  
August  
quality  
support  
protection  
pleasure

STORE

A workman needs good tools.  
He pushed his bicycle.  
Tomorrow there may be rainy weather.  
Not a soul was around.  
Flour is a white powder.  
Blood carries energy to our muscles.  
One end of your arm is at the shoulder.  
Sometimes a lawyer goes to court.  
January is the first month of the year.  
If you are on a diet you eat special food.  
Could you lift 100 pounds?  
That board is 12 feet long.  
The dog chased our cat in a circle.  
Boston is one of the largest cities in America.  
Some people are nervous if they have to make a speech.  
The road was not very smooth.  
Do you know who itches for the Red Sox?  
A woolen shirt is very warm.  
Our group finished first.  
I tried to get to school on time.  
Do you need to borrow a pencil?  
Two baseball games is called a double header.  
Summer is the vacation season.  
I just bought a new bicycle.  
The captain is in charge of this ship.  
Rabbits like to eat lettuce.  
Washington is the highest mountain in New England.  
He taught his brother to play baseball.  
A good ship always has to float.  
August is a summer month.  
Good quality tools will last a long time.  
Use this wood to support the shelf.  
Firemen and policemen work for our protection.  
It's always a pleasure to go on a trip.